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

CARACTERIZAÇÃO CONFORMACIONAL DE CARBOIDRATOS E GLICOPROTEÍNAS: EFEITOS DA GLICOSILAÇÃO NA GLICOPROTEÍNA α1-ÁCIDA HUMANA

Cláudia Lemelle Fernandes

Porto Alegre – Brasil
Fevereiro de 2011

CARACTERIZAÇÃO CONFORMACIONAL DE CARBOIDRATOS E GLICOPROTEÍNAS: EFEITO DA GLICOSILAÇÃO NA GLICOPROTEÍNA α1-ÁCIDA HUMANA

Cláudia Lemelle Fernandes

Tese de doutorado elaborada no Laboratório 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
Fevereiro de 2011

CARACTERIZAÇÃO CONFORMACIONAL DE CARBOIDRATOS E GLICOPROTEÍNAS: EFEITO DA GLICOSILAÇÃO NA GLICOPROTEÍNA α1-ÁCIDA HUMANA

Cláudia Lemelle Fernandes

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 Mestre em Biologia Celular e Molecular.

Banca Examinadora:

_____

Hugo Verli (Centro de Biotecnologia - UFRGS) (Presidente)

_____

Jorge Almeida Guimarães (Centro de Biotecnologia - UFRGS)

_____

Richard Charles Garett (Instituto de Física – USP-São Carlos)

_____

Roberto Lins (Departamento de Química Fundamental - UFPE)

_____

Charley C. Staats (Centro de Biotecnologia - UFRGS) (Suplente)

Esta tese foi realizada sob a orientação do professor Doutor Hugo Verli, com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) 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

FERNANDES, Cláudia Lemelle.

Caracterização conformacional de carboidratos e glicoproteínas: efeito da glicosilação na glicoproteína α1-ácida humana

Rio Grande do Sul, UFRGS, Centro de Biotecnologia, 2009.

Dissertação: Mestre em Ciências (Biologia Celular e Molecular)

1. Dinâmica Molecular 2. Glicoproteínas

3. Polissacarídeos 4. Ligação Glicosídica

I. Hugo Verli

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

III. Títulos

AGRADECIMENTOS

Meus sinceros agradecimentos ao Professor Hugo, pela sua importância na minha formação profissional e pelo exemplo de conduta científica.

Aos membros da comissão de acompanhamento, Dr. Rafael Roesler e, especialmente, Dra. Evelyn Koeche Schroeder, que além do suporte científico foi sempre uma amiga.

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

À Sílvia e ao Luciano, pelo bom humor e competência, mantidos em todas as situações.

Aos membros da banca examinadora, por aceitarem o convite.

Aos alunos do Grupo de Bioinformática Estrutural de maneira especial a Laércio Pol Fachin e Guilherme Menegon Giesel que me acolheram no início deste projeto, mas estendidos a todos pelo excelente ambiente de trabalho.

Ao Rodrigo Ligabue Braun, que mais do que colega se tornou um amigo para todas as horas.

Aos amigos, que mesmo sem entender exatamente o que fazemos no laboratório mantém seu apoio e compreendem as ausências, meu sincero obrigada.

Aos meus pais Valmir Dasso Fernandes e Nancy Maria Lemelle Fernandes, pelo apoio incondicional em todos os momentos e apostas, extendido a meu irmão, cunhada e sobrinhas pelo suporte familiar tão importante pra manter o equilíbrio nos momentos difíceis.

A CAPES e CNPq que financiaram este projeto.

"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale."

Marie Curie (1867-1934)

SUMÁRIO

LISTA DE ABREVIATURAS

AGP – Glicoproteína α_1-ácida

Ala - alanina

Asp – asparagina

Cox – ciclooxigenase

Cys – cisteína

DM – dinâmica molecular

Fuc – fucose

Gal – galactose

Glc – glicose

GlcNAc – N-acetil-glicosamina

Gly -– glicina

hCG – gonadotrofina coriônica humana

HSEA – esfera rígida exo-anomérica

IUPAC – *International Union of Pure and Applied Chemistry*

Hyl – hidroxilisina

Hyp – hidroxiprolina

IdoA – ácido idurônico

Leu – leucina

Man – manose

MD – *molecular dynamics*

MHC – complexo principal de histocompatibilidade

MM – mecânica molecular

NeuAc – ácido siálico

NMR – *nuclear magnetic resonance*

NOE – *nuclear Overhauser effect*

PDB – *Protein Data Bank*

Phe – fenilalanina

Pro – prolina

PME – *Particle-Mesh Ewald*

RE – retículo endoplasmático

RMN – ressonância magnética nuclear

Ser – serina

SleX – oligossacarídeo *Sialil Lewis x*

Thr – treonina

Tyr – tirosina

Xyl – xilose.

RESUMO

O conhecimento da glicobiologia estrutural se mantém como a parte menos explorada do estudo de estrutras tridimensionais. Considerando que a glicosilação pode estar envolvida em processos biológicos como crescimento celular e inflamação o descrever das bases moleculares da interação proteína-carboidrato pode auxiliar na compreensão destes eventos. Neste sentido considerando o pequeno número de trabalhos avaliando o perfil conformacional de glicoproteínas por simulação de dinâmica molecular e RMN este trabalho demonstrou que o campo de força GROMOS43a1 é capaz de representar adequadamente um conjunto de glicoproteínas tendo como conformação inicial as estrutras determinadas por RMN. O próximo desafio foi simular dissacarídeos isolados em solução e comparar o seu perfil aos estudos anteriores, o que demonstrou que o conjunto de conformações de cada dissacarídeo representa as conformações obtidas em ambos os métodos. Para validar o uso de conformações de dissacarídeos como unidades de construção de glicanas em glicoproteínas foi descrita a forma glicosilada das enzimas Cox-1 e Cox-2, que não possuíam a estrutura da glicosilação, e as conformações das glicanas nestas simulações foram similares as dos dissacarídeos, comprovando que as conformações de dissacarídeos podem ser extrapoláveis para construção de glicoproteínas. Finalmente foi construída a forma glicosilada da AGP humana, uma proteína que possui diferenças funcionais associadas ao perfil das glicanas ligadas. Sabe-se que em casos de resposta inflamatória a concentração plasmática da AGP pode ser aumentada em até cinco vezes a concentração normal e durante este aumento há uma diferença no número de ramificações das glicanas e presença de resíduos de fucose. Foram simuladas três estrutras AGP, sem glicosilação e glicosilada com a presença e ausência de fucoses,. A flexibilidade da estrutura não glicosilada é muito maior que das glicoproteínas, mostrando o papel estrutural da glicosilação. Adicionalmente foi estudada a interação que a AGP com selectinas, proteínas envolvidas nos processos inflamatórios, fornecendo dados preeliminares do papel molecular da interação AGP-selectina.

ABSTRACT

The knowning of strucutural glycobiology still the less explored area in threedimensional structures. Considering the envolviment of glycosylation in biological process such cellular grown and inflammation describe the molecular basis interactions of protein-carbohidrate may facilitate understanding of these events. In this way considering the small number of works evaluated the conformational profile of glycoproteins by molecular dynamics simulations and NMR this work demonstrate that the GROMOS96 43a1 force field adequately represent a glycoprotein's conformational ensemble taking as the starting geometries, the NMR determined structures. The next step is simulate isolated disaccharides in solution and compare these proflite with previous work, which demonstrate that the conformational ensamble of disaccharides represents the conformations obtain in both methods. To validate the use of disaccharides conformations such construct units of glycans in glycoproteins, the glycosylated form of Cox-1 and Cox-2 with no previous structure were simulated, and the conformations of glycans were similar to disaccharides, proving that disaccharides conformations can be extrapolated to construct glycoproteins. Finaly it was construct the glycosylated form of human AGP, a protein with functional differentces associated to glycan linked profile. In cases of inflammatory response the the plasma concentration can rise up to fivefold and in this case the glycans differ in the branching and presence of fucose residues. Three structures of AGP were construct, unglycosylated and glycosylated with and without fucoses. The structural flexibility of unglycosylated form was higher than the glycosylated forms, demonstrated the structural role of glycans. Additionally it was study the interaction of AGP with selectins, proteins envolved in inflammatory process, supply preliminary data to molecular role of AGP-selectin interaction.

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

1 Introdução

1.1 Glicosilação de proteínas

Glicoproteínas são biomoléculas envolvidas em uma grande variedade de processos biológicos, incluindo fertilização, defesa imunológica, replicação viral, infecções parasitárias, crescimento celular, adesão celular, degradação de coágulos e inflamação, dentre outros (Dwek, 1996). São constituídas por uma parte protéica ligada a uma porção sacarídica, as quais se associam no ambiente celular em um processo denominado glicosilação. Este processo pode ocorrer tanto pós- como co-traducionalmente em todos os organismos conhecidos, sendo controlada por fatores diversos dependentes tanto do tipo celular envolvido como da espécie em questão (Spiro, 2002).

A adição de carboidratos à estrutura protéica envolve uma grande diversidade de possíveis monossacarídeos ligados a diversos resíduos de aminoácidos. As conexões mais comumente encontradas em glicoproteínas incluem a N-glicosilação e a O-glicosilação, identificadas em eucariotos, árqueas e eubactérias (Spiro, 2002; Schmidt *et al.*, 2003; Abu-Qarn *et al.*, 2008), embora sistemas como a C-glicosilação, fosfoglicosilação e glipiação tenham sido descritos. Ao todo, até o momento foram identificados cinco tipos de glicosilação, envolvendo treze monossacarídeos e oito resíduos de aminoácidos diferentes (Dell e Moris, 2001; Spiro, 2002; Pol-Fachin & Verli, 2011). Adicionalmente, mesmo para uma dada proteína a glicosilação é um tipo de modificação na estrutura peptídica altamente variável, uma vez que uma mesma proteína pode apresentar diferentes tipos de glicosilação (Medvedová & Farkas, 2004).

A síntese de oligossacarídeos na N-glicosilação corresponde ao tipo de conexão proteína-carboidrato mais amplamente estudado, tendo parte de suas vias biossintéticas e etapas envolvidas já caracterizadas (Helenius & Aebi, 2001). Envolve dois possíveis resíduos de aminoácidos, Asn (mais comum) e Arg. Particularmente no caso do resíduo Asn, a glicosilação é determinada por uma sequência consenso de quatro resíduos de aminoácidos, a saber (Marshall, 1972):

N {P} S/T {P}

- Na primeira posição um resíduo de Asn;
- Na segunda posição qualquer aminoácido que não Pro;
- Na terceira posição um resíduo de Ser ou Thr;
- Na quarta posição qualquer aminoácido que não Pro.

Em linhas gerais, o processo de N-glicosilação pode ser descrito por uma série de dez passos (Figura 1) (Helenius & Aebi, 2001):

a. Na face citosólica da membrana do retículo endoplasmático (RE) é sintetizado um heptassacarídeo (composto por dois resíduos de GlcNAc e cinco de Man), conectado a uma molécula de dolicol fosfato;

b. Este sistema heptassacarídeo-dolicol é translocado para o lúmen do RE;

c. O heptassacarídeo inicial sofre a adição de mais sete resíduos de monossacarídeos, incluindo Man e Glc;

d. O oligossacarídeo então formado é transferido para a cadeia polipeptídica nascente, de forma co-traducional, ou ainda de forma pós-traducional;

e/f. A glicoproteína então recém formada sofre ciclos de adição e remoção de resíduos de Glc;

g. A partir da estrutura sacarídica madura, conclui-se o enovelamento correto da estrutura protéica;

h. A glicoproteína pode ser transferida para o complexo de Golgi para posterior processamento ou, em outros casos, ser direcionada para seu local de ação. Caso a mesma não passe pelo Golgi, a estrutura sacarídica da glicoproteína será aquela determinada pelo sistema oligossacarídico originado junto ao dolicol fosfato;

i. Remoção de resíduos de Man da(s) cadeia(s) de glicana(s) ligada(s) à proteína;

j. E etapas posteriores podem envolver ciclos de adição de outros resíduos para formação de diferentes estruturas oligossacarídicas, tais como GlcNAc, Gal e NeuAc;

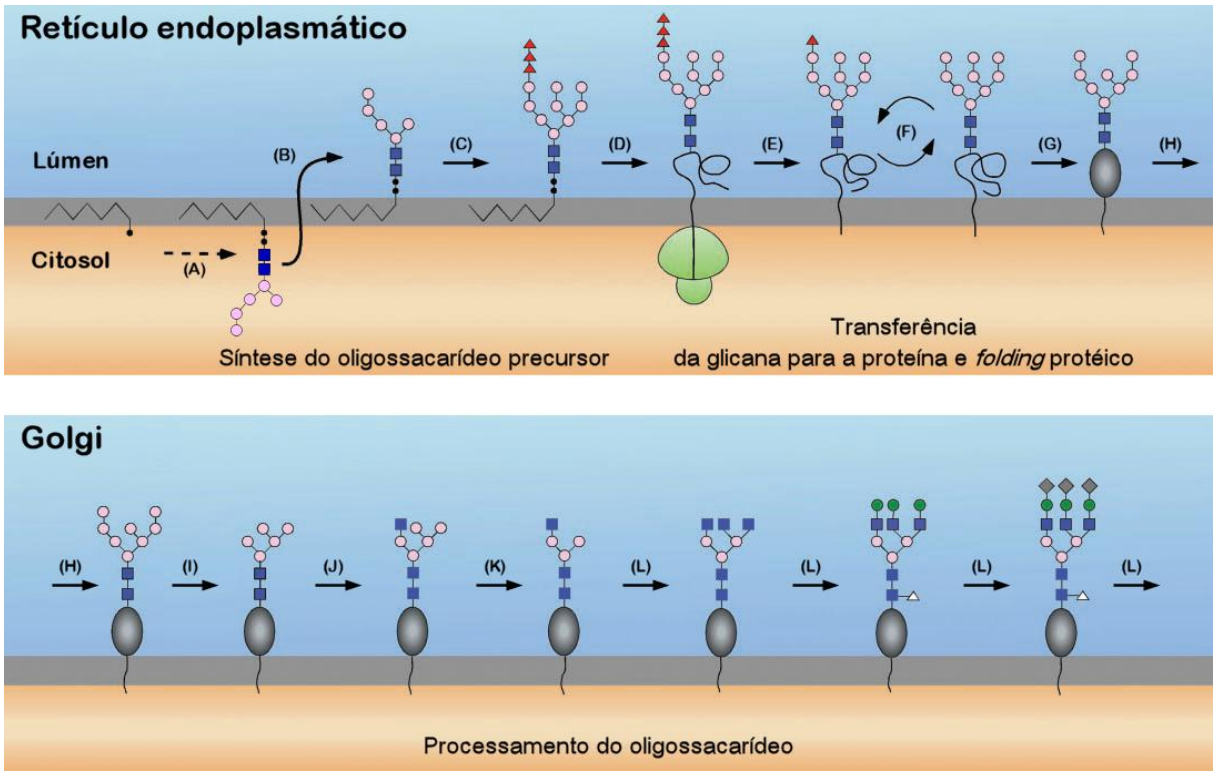


Figura 1: Processo de biossíntese de um oligossacarídeo N-ligado do tipo complexo. As formas azuis representam resíduos de GlcNAc; as rosas, Man; as vermelhas, Glc; as verdes, Gal; as cinzas, NeuAc; e as brancas, Fuc. (Helenius & Aebi, 2001, adaptado de Pol-Fachin, 2009).

Uma vez alcançado o complexo de Golgi, a estrutura sacarídica da glicoproteína recém sintetizada pode ser ainda convertida em diferentes tipos de glicanas, dependendo da proteína a que esse carboidrato está ligado e de fatores teciduais celulares e metabólicos, (Spiro, 2002). Estes oligossacarídeos sintetizados podem ser classificados em três grupos principais: I) oligomanose (ou *high manose*), quanto todos os resíduos de carboidrato substituintes são Man; II) complexo, quando estes resíduos são de qualquer outro monossacarídeo; e III) híbridos, quando parte das ramificações mantém somente resíduos de Man e parte tem inseridos outros tipos de monossacarídeos (Figura 2) (Dwek, 1996).

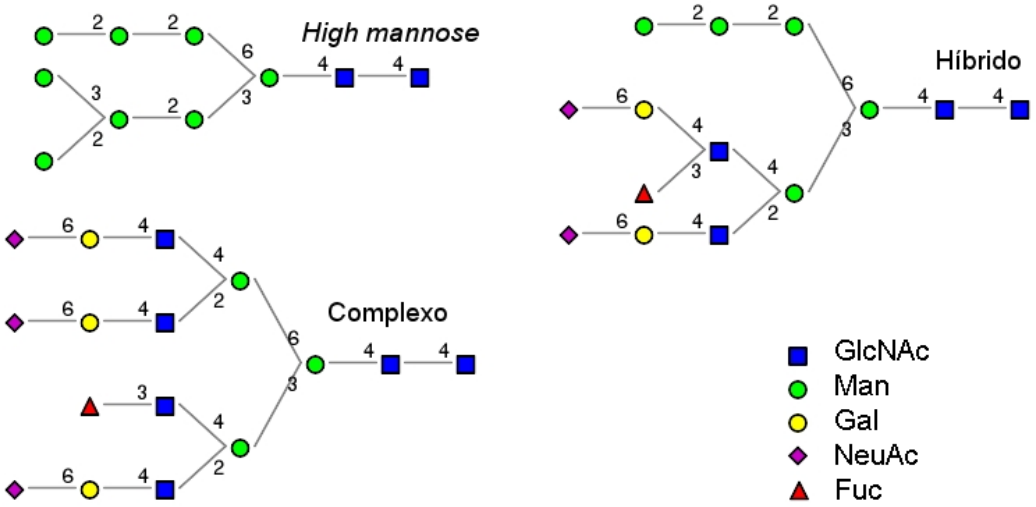


Figura 2: Tipos de oligossacarídeos que podem ser encontrados ligados em N-glicosilação.

Resultados obtidos em modelos animais encontraram evidências de que os tipos de oligossacarídeos, ligados a uma dada proteína, podem mudar durante o desenvolvimento do organismo, assim como tipos específicos de glicanas podem ser expressos em diferentes estágios da diferenciação e até ser expressa mais de um tipo de glicosilação simultameamente para a mesma proteína (Cipollo *et al.*, 2005). Tal miríade de padrões de glicosilação sugere uma enorme plasticidade estrutural e funcional para glicoproteínas (Fukuda, 1991; Saéz *et al.*, 2001), o que pode estar relacionado à sua presença ubíqua nos organismos – estima-se que 70% das proteínas possuam sítios para N-glicosilação (Varki, 2002) e que mais de 50% de todas as proteínas na natureza sejam glicosiladas (Lutteke *et al.*, 2006). Exemplos da plasticidade de funções de glicanas a partir de mudanças na sua composição sacarídica podem ser encontradas em patologias como câncer (Couldrey & Green, 2000; Gunnarsson *et al.*, 2007; Takahashi *et al.*, 2008) e um grupo de patologias derivadas de defeitos no processo de N-glicosilação, ao que se denomina síndrome de glicoproteínas com deficiência de carboidratos (Jaeken & Carchon, 1993).

Em comparação à N-glicosilação, a ligação O-glicosídica apresenta maior variabilidade considerando-se a elevada quantidade de aminoácidos envolvidos, *i.e.* Ser, Thr, Tyr, Hyp e Hyl. Esta variedade pode estar relacionada à dificuldade encontrada até o momento no estabelecimento de seqüências consenso associadas à O-glicosilação (Spiro, 2002). No entanto, o desenvolvimento de bases de dados e

algoritmos de predição (Gupta *et al.*, 1999; Li *et al.*, 2006) possibilitaram a identificação de algumas propriedades das regiões de O-glicosilação, tais como a preferência por aminoácidos como Pro e aversão por aminoácidos aromáticos nas regiões adjacentes à ligação proteína-carboidrato (Thanka Christlet & Veluraja, 2001).

A síntese da O-glicosilação é atribuída a vias diferentes entre eucariotos e eubactérias (que não são capazes de realizar N-glicosilação). Em eubactérias, a ligação de carboidratos a proteínas ocorreria através da transferência pós-traducional de monossacarídeos isolados como GalNAc, Man, Xyl e Fuc adicionadas a aminoácidos que possuam um grupamento hidroxila livre em polipeptídeos já enovelados (Imperiali & Hendrickson, 1995).

Tendo sido descritos apenas recentemente, e somente em eucariotos, pouca informação encontra-se disponível para fosfoglicosilação e C-glicosilação. Em linhas gerais, a fosfoglicosilação se caracteriza pela ligação de oligossacarídeos a resíduos de Ser ou Thr via ligações fosfodiéster (Mehta *et al.*, 1996; Haynes, 1998). Já a C-glicosilação ou C-manosilação é a ligação de um resíduo de α-Man adicionado ao grupamento indol de um Trp presente na primeira posição da seqüência consenso Trp-X1-X2-Trp/Cys. Essa ligação é favorecida quando X1 é um aminoácido pequeno ou polar, como Ser, Ala, Gly ou Thr, e desfavorecida quando X1 é Phe ou Leu (Hofsteenge *et al.*, 1994; de Beer *et al.*, 1995; Krieg *et al.*, 1998; Doucey *et al.*, 1998; Julenius, 2007).

Por fim a glipiação, identificada em eucariotos e árqueas (Spiro, 2002), envolve a ligação de GPI à região C-terminal de polipeptídeos visando ancorar a proteína, glicoproteína ou proteoglicana à superfícies celulares (Low, 1989; Paulick & Bertozzi, 2008).

1.2 Um caso de estudo: glicoproteína α_1-ácida (AGP) humana

A AGP humana é uma glicoproteína plasmática, também conhecida como orosomucoide, caracterizada inicialmente em 1950 como cristais isolados de uma fração plasmática (Schmid, 1950; Schmid, 1953) (Figura 3). Com relação às proteínas plasmáticas, a AGP só é menos abundante que a albumina. Dentre as suas características principais estão ser uma proteína incomum com um ponto isoelétrico muito baixo (entre 2,8 e 3,8) e um alto conteúdo de carboidratos (cerca de

45% de sua massa molecular), distribuídos em cinco sítios de glicosilação. Isto a torna umas das glicoproteínas conhecidas com maior teor de carboidratos em relação à sua massa total (Fournier *et al.* 2000). Embora a função da AGP não seja completamente entendida, ela é considerada um agente antiinflamatório e imunomodulatório natural, e esta atividade imunomodulatória é dependente de sua glicosilação (Fournier *et al.*, 2000).

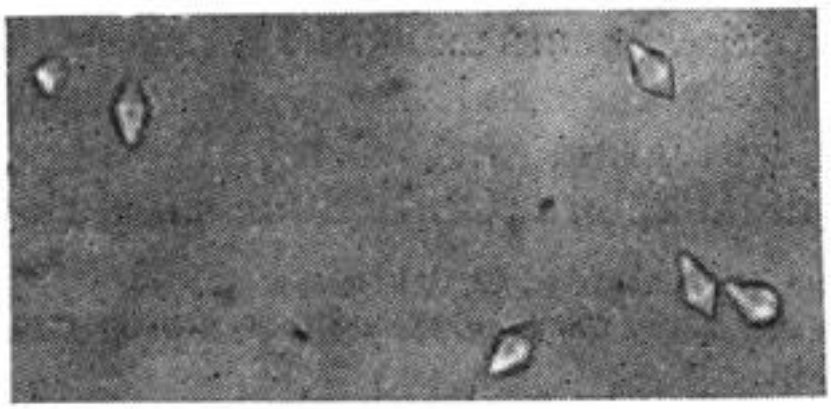


Figura 3: Cristais de AGP isolados em 1950 por Schmid.

A AGP circulante possui de 41 a 43 KDa, incluindo duas isoformas de 183 aminoácidos (apenas 22 aminoácidos diferem entre as duas variantes, ORM1 e ORM2). A variante ORM2 é expressa pelos genes *AGP-B* e *AGP-B'* e sua expressão é pelo menos cem vezes menor que o ORM1. Já a variante ORM1 é produto do gene *AGP-A*, expresso principalmente em hepatócitos e o único responsável pela resposta de fase aguda. Sua expressão é regulada por várias citocinas pró-inflamatórias, como IL-6 e IL-12, e glicocorticóides. Entretanto, existem relatos de síntese extra-hepática da AGP, relacionada principalmente à respostas de fase aguda (Fournier *et al.* 2000). De fato, sua concentração plasmática pode estar aumentada em até cinco vezes em processos inflamatórios (Schonfeld *et al.*, 2008), o que a torna a proteína majoritária na resposta inflamatória de fase aguda (Fournier, *et al.*, 2000).

Estruturalmente, a AGP possui duas pontes dissulfeto entre os resíduos Cys5-Cys147 e Cys72-Cys165 (Schmid *et al.*, 1974). Seu enovelamento foi proposto como do tipo lipocalina a partir da comparação de sua seqüência com outras estruturas homólogas ou similares e de estudos de dicroísmo circular (Rojo-Rodriguez & Hernandez-Aranha, 1993, Kodicek *et al.*, 1995). Esta proposta foi posteriormente

confirmada com a determinação da estrutura tridimensional da AGP humana por cristalografia de raios-X (Schonfeld *et al.*, 2008).

Os dados cristalográficos indicaram que a AGP apresenta oito α-hélices e nove fitas-β antiparalelas, fitas estas que formam a cavidade central da proteína, definindo um barril-β. Neste barril forma-se uma cavidade envolvida na ligação e carreamento de fármacos (Wassan *et al.*, 2008) (Figura 4).

Juntamente com a albumina e lipoproteínas, a AGP é um dos mais importantes transportadores de fármacos no plasma. Pode se ligar a uma grande variedade de compostos, preferencialmente de natureza básica ou neutra, dado seu baixo ponto isoelétrico, o que faz uma contrapartida à albumina, que liga-se a compostos ácidos (Kremer, 1988). Sua importância na ligação a fármacos faz da AGP uma proteína essencial para o estudo e desenvolvimento de novos agentes candidatos a fármacos (Fuse *et al.*, 1999, Parikh *et al.*, 2000).

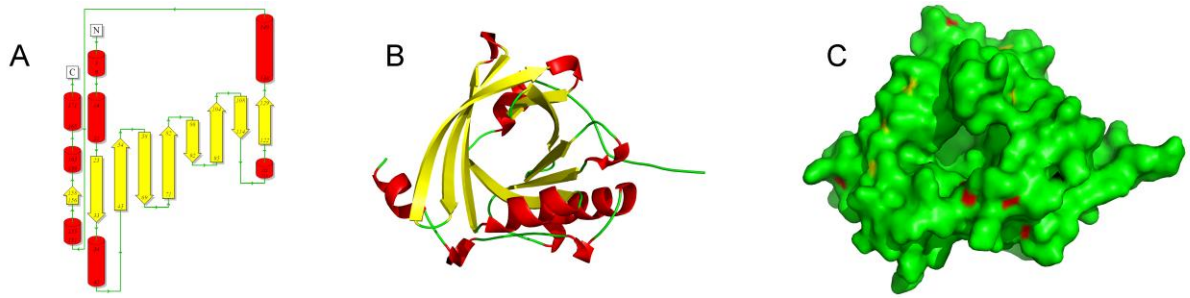


Figura 4: A) Representação esquemática da estrutura secundária da AGP evidenciando o arranjo de fitas-β antiparalelas (amarelo) que caracteriza o enovelamento do tipo lipocalina. A estrutura tridimensional da AGP em duas representações com a mesma orientação, hélices e fitas (B) e superfície (C) evidenciando que este arranjo de fitas forma uma cavidade central na proteína, envolvido na ligação a fármacos.

1.2.1 A estrutura de glicosilação da AGP

Considerando-se o grande espectro de funções apresentado pela AGP, assim como o fato de que quase metade de sua massa é composta por carboidratos, pode-se considerar que a elucidação de sua estrutura de glicosilação poderá oferecer importantes contribuições para o entendimento de sua biologia estrutural e função biológica. De acordo, é relatado que a adição de carboidratos a uma dada estrutura

protéica é capaz de influenciar diversas de suas propriedades, tais como resistência a desnaturação, aumento de solubilidade, prevenir agregação e alterar o reconhecimento entre proteínas e de proteínas com carboidratos (Varki, 1993; Skropeta, 2009).

Diversos estudos (Stubbs *et al.*, 1997; Fournier *et al.*, 2000; Sei *et al.*, 2002; Nakano *et al.*, 2003; Nagy *et al.*, 2004), principalmente aqueles empregando espectrometria de massas, vêm se dedicando à identificar as glicanas mais comumente ligadas na AGP. Assim, algumas características foram associadas aos oligossacarídeos de posições específicas desta glicoproteína (Treuheit *et al.*, 1992; Fournier *et al.*, 2000):

- O sitio de glicosilação no resíduo Asn15 nunca carrega um oligossacarídeo tetra-ramificado e geralmente carrega um oligossacarídeo tri-ramificado;
- O sitio de glicosilação no resíduo Asn38 nunca carrega oligossacarídeos com fucose, e geralmente carrega oligossacarídeos bi-ramificados;
- O sitio de glicosilação no resíduo Asn75 nunca carrega um oligossacarídeo bi-ramificado;
- Os sítios de glicosilação nos resíduos Asn54, Asn75 e Asn85 apresentam geralmente oligossacarídeos tetra-ramificados;

A despeito destes trabalhos, a estrutura de glicosilação exata da AGP, incluindo a abundância relativa do tipo de glicosilação associado a cada posição de N-glicosilação na AGP, ainda não foi relatada. Contudo, dados anteriores apresentam os oligossacarídeos mais abundantemente ligados a AGP (Figura 5). Entretanto as ligações dos NeuAc com as Gal, que são os resíduos mais comumente encontrados nas terminações das glicanas associadas a AGP, não puderam ser identificados (Fournier *et al.*, 2000; Sei *et al.*, 2002).

A determinação da estrutura sacarídica da AGP encontra ainda dificuldades adicionais. Processos inflamatórios não somente podem interferir no nível de expressão da AGP como também na sua composição. Grupos de pesquisa vêm demonstrando que durante o choque séptico o grau de ramificação e o conteúdo de resíduos de fucose são aumentados (Graaf, *et al.*, 1993; Linden *et al.*, 1996; Olewicz-Gawlik *et al.* 2007; Levander, *et al.*, 2009).

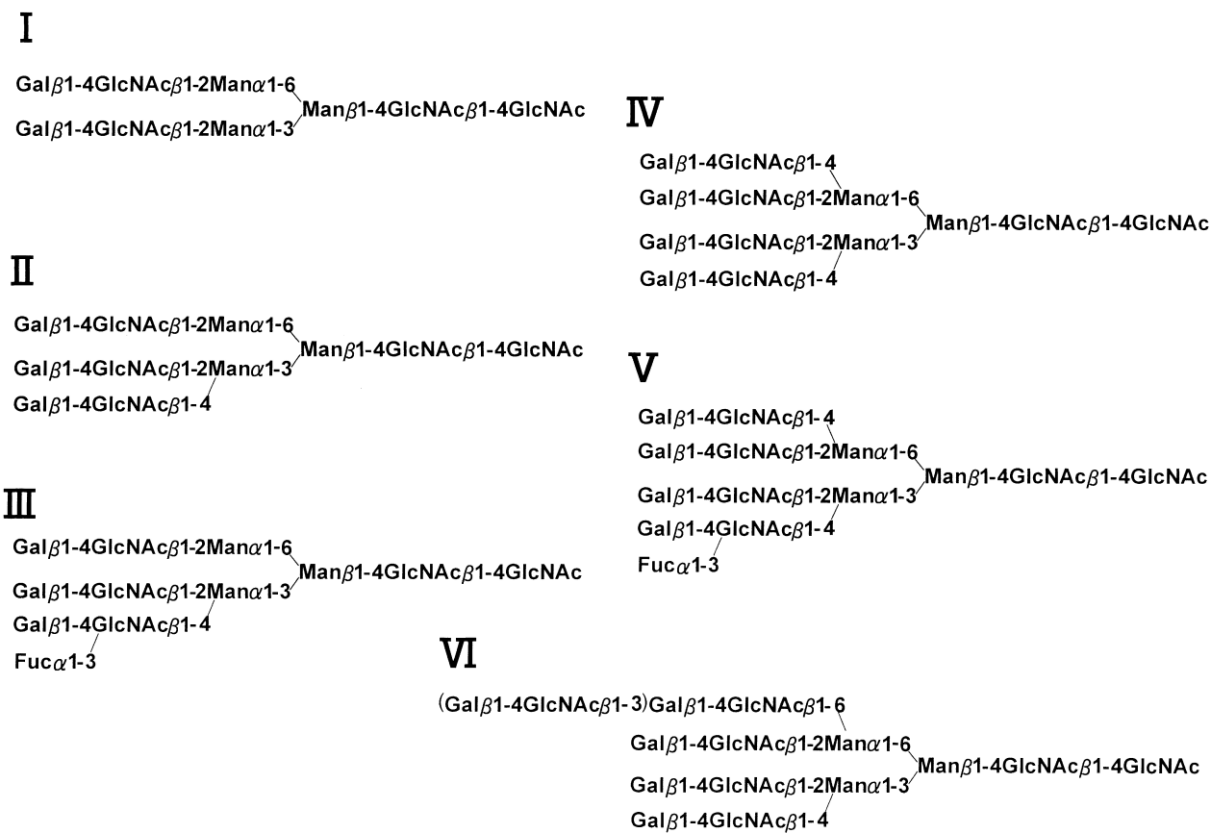


Figura 5: Oligossacarídeos mais abundantes encontrados ligados AGP. Adaptado de Sei *et al.*, 2002.

Esta adição de resíduos de fucose tem uma importante implicação funcional. Ao ser adicionado às árvores de carboidratos da AGP, em posições específicas, este resíduo origina um epítopo tetrassacarídico denominado Sialil Lewis x (SLeX) (Fournier, *et al.*, 2000) (Figura 6).

Este grupamento é relatado na literatura como envolvido em processos de reconhecimento de uma classe de lectinas com função imunomodulatória, as selectinas (Foxall, *et al.*, 1992; Laski, 1992; Somers *et al.*, 2000). Selectinas são uma família de glicoproteínas de superfície celular responsáveis por eventos de adesão no reconhecimento de leucócitos em sítios de inflamação e a migração destes para o tecido linfático. Existem três tipos de selectinas, a E-selectina e a P-selectina são induzidas na superfície do endotélio vascular como resposta ao estímulo inflamatório, mas com diferentes relações de expressão cinética. A L-selectina é expressa constitutivamente nos leucócitos com a função de amplificar o processo inflamatório promovendo interações entre leucócitos ou entre leucócitos e endotélio (Kansas, 1996; Vestweber & Blanks, 1999; Somers *et al.*, 2000).

O SLeX é o principal ligante fisiológico das moléculas de células de adesão como a E-selectina e a P-selectina, envolvidas na adesão de neutrófilos, monócitos, ou células T em repouso, ou ainda plaquetas durante a inflamação (Lasky, 1992).

Desta forma, a inserção de resíduos de fucose nas glicanas da AGP, dando origem ao grupamento SLeX durante o processo inflamatório torna possível a ligação da AGP tanto à E-selectina como com a P-selectina endoteliais, competindo com os leucócitos que expressam naturalmente o ligante das selectinas, o SLeX (Fournier *et al.*, 2000). Consequentemente, esta ligação pode representar um mecanismo de inibição do extravasamento leucocitário no tecido inflamado uma vez que a AGP bloquearia a ligação da selectina com o leucócito (Fournier *et al.*, 2000).

Conforme dados anteriores, o SLeX está presente apenas nos oligossacarídeos tri ou tetra ramificados da AGP (Fournier *et al.*, 2000) e somente nas posições Asn54, Asn75 e Asn85, o que justifica também o aumento do grau de ramificação em casos de choque séptico. Neste contexto, trabalhos posteriores relacionam o aumento das ramificações e formação de SLeX na AGP a casos de câncer e outras doenças que possuam reações de fase aguda (Fournier *et al.*, 2000, Gunnarsson *et al.*, 2007). Como exemplo, pode-se mencionar o estudo de Imre e colaboradores onde o perfil de glicosilação da AGP de pacientes saudáveis é claramente diferenciado de pacientes com câncer (Imre *et al.*, 2008)

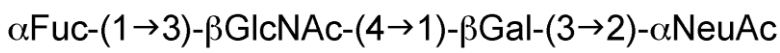


Figura 6: Tetrassacarídeo do tipo SLeX. Este fragmento sacarídico é encontrado frequentemente ligado a oligossacarídicos da AGP mas somente nas posições Asn54, Asn75 e Asn85 e principalmente em casos de aumento da concentração de AGP plasmática em inflamações, respostas agudas.

1.3 Estrutura e conformação de proteínas

O conhecimento de estruturas tridimensionais de proteínas tem um papel extremamente importante dentro da biologia molecular. O papel de destaque dos esforços de caracterização da estrutura tridimensional de proteínas pode ser evidenciado no número crescente de estruturas depositadas no PDB (Tabela 1) (Berman *et al.*, 2000). Neste processo, dois métodos apresentam importância e papel destacados: a cristalografia de raios X e a RMN.

Tabela 1: Dados de Estruturas depositadas no PDB[a]

Método experimental	Proteínas	Ácidos nucleicos	Complexos Prot/AcNuc	Outros	Total
Raio X	57073	1252	2736	17	61078
RMN	7599	933	167	7	8706
Microscopia eletrônica	233	22	85	0	340
Híbrido	26	1	1	1	29
Outro	129	4	4	13	150
Total	65060	2212	2993	38	70303

[a]De acordo com «http://www.pdb.org/pdb/statistics/holdings.do» em 7JAN2011.

Conforme pode ser observado na Tabela 1, a cristalografia de raios X constitui-se na principal fonte de novas estruturas de proteínas. Esta é a principal técnica para a obtenção de estruturas tridimensionais de proteínas e, em linhas gerais, está baseada na obtenção de cristais regulares que ao serem submetidos a uma fonte de raios X produzindo perfis de dispersão diferente em cada parte do cristal. Esta amplitude de radiação dispersa por um ponto é proporcional a densidade eletrônica neste ponto (Leach, 2001). A cristalização de proteínas, entretanto, é um processo lento, complexo e, algumas vezes, impossível de realizar-se. As dificuldades abrangem tanto o processo de produção e purificação da proteína de interesse quanto a indução da cristalização e o transporte dos cristais (Branden & Tooze, 1999).

Entretanto pela necessidade de obtenção de cristais, as condições empregadas na tentativa de sua obtenção costumam envolver ambientes de hipersaturação, onde as unidades proteicas ficam demasiadamente próximas umas das outras, ocasionando contatos não biológicos que podem, em algumas situações, modificar a conformação da proteína (Carugo & Argos, 1997; Eval *et al.*, 2005). As regiões mais suscetíveis a essas alterações são, geralmente as, áreas da superfície da proteína, como alças, por apresentarem maior flexibilidade e menor ancoramento a outros elementos de estrutura secundária (Copié *et al.*, 1998; Andrec *et al.*, 2007).

Em contrapartida, a resolução da estrutura tridimensional de proteínas através da técnica de RMN baseia-se na resolução de sinais de NOESY de átomos

(geralmente hidrogênios) que estão próximos no espaço mas que podem estar separados por muitas ligações químicas (Leach, 2001). Em proteínas devido a quantidade de átomos envolvidos a técnica se torna complexa, sendo aplicada geralmente a proteínas com aproximadamente 35 KDa, entretanto fato das proteínas estarem em solução e a possibilidade de produzir diversas conformações para uma mesma solução torna a técnica adequada para descrever sistemas em solução (Branden & Tooze, 1999).

A obtenção de dados pela técnica de RMN em solução traz algumas vantagens em relação à cristalografia como, por exemplo, a possibilidade de se empregar condições com concentrações de proteína mais próximas daquelas observadas fisiologicamente evitando, assim, efeitos conformacionais devido ao empacotamento cristalino, assim como a obtenção de uma variedade de conformações e não uma estrutura média das conformações (Branden & Tooze, 1999). Entretanto estudos de cristalografia onde mais de um modelo cristalográfico é resolvido para uma mesma estrutura podem auxiliar no entendimento de propriedades dinâmicas de proteínas em regiões de alta flexibilidade (Navarro, *et al.*, 2007).

1.4 Estrutura e conformação de carboidratos

Carboidratos são moléculas biológicas compostas principalmente por carbono, oxigênio e hidrogênio abundantes na natureza. Eles podem ser encontradas em vários componentes da dieta animal, em materiais estruturais de plantas (celulose) e em peptídeoglicanos de bactérias, além de funções biológicas como sinais de reconhecimento de superfície celular para anticorpos, hormônios, toxinas entre outros (Rao, *et al.*, 1998).

A unidade básica de uma molécula de carboidrato é conhecida como monossacarídeo, representada pela fórmula geral $C_nH_{2n}O_n$. Sendo o número de carbonos variável entre três unidades (treoses) até seis unidades (hexoses). Uma das características marcantes de carboidratos é a presença de centros quirais nos carbonos, fazendo que os mesmos tenham tanto isômeros como enantiômeros. Em solução, carboidratos com mais de quatro carbonos geralmente estão na forma cíclica, que cria mais um carbono quiral na posição um da molécula sacarídica, conhecido como carbono anomérico (Rao, *et al.*, 1998).

Dentre os carboidratos mais abundantes encontrados temos os monossacarídeos cíclicos de pentoses (furanoses) e hexoses (piranoses) que, assim como aminoácidos e ácidos nucléicos, são unidades de formação de cadeias, os polissacarídeos. Entretanto na formação de cadeias oligossacarídicas carboidratos se diferenciam de outras classes de macromoléculas por duas importantes características: I) eles podem ser altamente ramificados, e II) suas unidades anoméricas podem se ligar umas as outras de diferentes formas. Estas características dão aos carboidratos uma grande variedade de estruturas possíveis, o que aumenta muito a complexidade do estudo conformacional e estrutural destas moléculas (Varki, 1993; Dwek, 1996). A ligação de carboidratos a proteínas pode acontecer com oito aminoácidos diferentes e treze diferentes tipos de dissacarídeos, o que devido ao carbono anomérico onde é feita a ligação produz 31 combinações diferentes só de ligação de carboidratos a proteínas (Spiro, 2002). Além disso aminoácidos a ácidos nucléicos se ligam somente de uma forma uns aos outros enquanto carboidratos podem estar ligados em quatro carbonos diferentes (1→2, 3, 4 ou 6) (Spiro, 2002).

A conformação de carboidratos é usualmente descrita de acordo com as recomendações da IUPAC (IUPAC-IUB, 1980; IUPAC-IUB, 1983), segundo as quais o arranjo espacial dos átomos de um monossacarídeo piranosídico, em sua forma cíclica, é determinado por uma letra maiúscula, em itálico, designando a forma do anel, e por números, responsáveis pela distinção das possíveis variantes conformacionais para um mesmo resíduo, como nas formas de cadeira $^{4}C_{1}$ e $^{1}C_{4}$ (Figura 7A e 7B). Adicionalmente, monossacarídeos podem assumir os estados configuracionais dextrógiro (D-) ou levógiro (L-), que diferenciam-se pela organização espacial de seus substituintes (Figura 7C e 7D), e podem ser encontrados, ainda, nos estados anoméricos α e β, no que tange à posição axial ou equatorial de seu substituinte na posição C1 (Figura 7D e 7E).

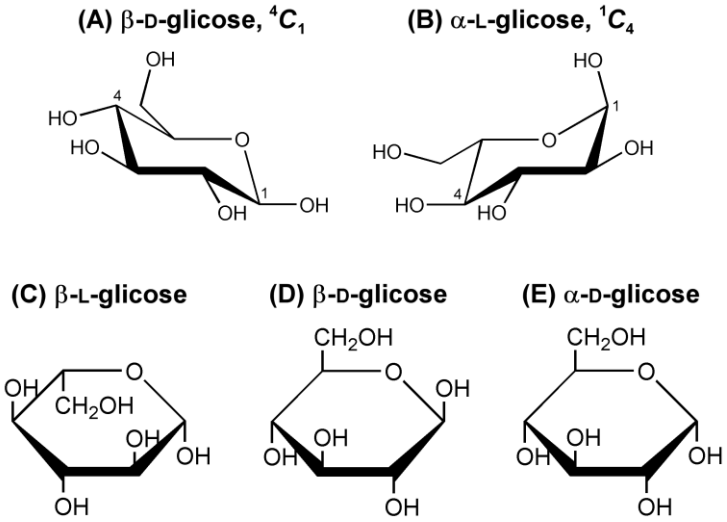


Figura 7: Variantes conformacionais e configuracionais associados a monossacarídeos: conformações do anel piranosídico (A e B), estados levógiro (C) ou dextrógiro (D) e estado anomérico α ou β (D e E). Nessas representações, a glicose é usada como modelo.

Como resultado das propriedades tanto das unidades monossacarídicas quanto das suas interconexões, carboidratos se apresentam como moléculas extremamente flexíveis, com múltiplos estados conformacionais simultaneamente co-existindo em solução (Spiro, 2002). Isto gera uma série de implicações para estudos buscando sua determinação estrutural, particularmente aqueles em nível atomístico. De fato, são bastante raras as estruturas tridimensionais de glicoproteínas depositadas no PDB, sejam elas obtidas por métodos de RMN (Tabela 2) ou cristalografia de raios X.

Por exemplo, no que concerne à metodologia de cristalografia de raios X, esta flexibilidade pode resultar em dificuldades na obtenção tanto de cristais quanto de densidades eletrônicas adequadas ao redor das estruturas sacarídicas comprometendo, assim, a determinação adequada de sua estrutura e conformação (Petrescu *et al.*, 1999). Dessa forma, não são incomuns erros na determinação do estado anomérico ou a obtenção conformações distorcidas para os monossacarídeos estudados (Lütteke *et al.*, 2004), ou ainda na orientação relativa de dois resíduos unidos através de uma ligação glicosídica (Lütteke, 2009). A dimensão deste problema pode ser evidenciada em um estudo sugerindo que aproximadamente um terço das estruturas depositadas no PDB contendo

carboidratos apresenta erros importantes com relação a estrutura glicídica (Crispin *et al.*, 2007; Berman *et al.*, 2007).

Tabela 2: Estruturas tridimensionais de glicoproteínas e polissacarídeos determinados por RMN.

Moléculas		Número de monossacarídeos[b]	Código PDB
Geral	Específica[a]		
Glicoproteína	α-hCG	9	1HD4
	Domínio de adesão – CD2	9	1GYA
	Calcitonina	8	1BZB
	Gliprot. N-ligada de *C. jejuni*	7	2K33
	Domínio extracelular – CD152	4	1AH1
	CD59	3	1CDR
	Fator de crescimento epidermal	2	2RQZ
	Conotoxina – *Conus textile*	2	1WCT
	Domínio EGF – fVIIa	1	1FF7
	Domínio lectina – Latrofilina	1	2JXA
	Domínio – Fibronectina	1	1E88
	Domínio – Fibronectina	1	2FN2
	Toxina – *Corynespora cassiicola*	1	2HGO
	Domínio EGF – Trombomodulina	1	1DQB
Polissacarídeo	Heparina	12	1HPN
	Ácido hialurônico	8	2BVK

[a] Referências, em ordem de aparecimento: Erbel *et al.*, 2000; Wyss *et al.*, 1995; Hashimoto *et al.*, 1999; Metzler *et al.*, 1997; Slynko *et al.*, 2009; Fletcher *et al.*, 1994; Hosoguchi *et al.*, 2010; Rigby *et al.*, 1999; Kao *et al.*, 1999; Vakonakis *et al.*, 2008; Pickford *et al.*, 2001; Sticht, *et al.*, 1998; Barthe *et al.*, 2007; Wood *et al.*, 2000; Mulloy *et al.*, 1993; Almond *et al.*, 2006;

[b] Número de monossacarídeos unidos em um único oligossacarídeo.

[c]Revisado em 15JAN2011.

Em se tratando da técnica de RMN, o desafio se concentra em resolver os sinais de NOE tanto devido a flexibilidade dos glicídeos, produzindo mais sinais de

NOE intra-resíduo do que inter-resíduo. A flexibilidade das glicanas ainda produz mais um problema na determinação do RMN: a obtenção do sinal de NOE é realizada numa escala de tempo entre 50 ms e 1s o que, pra sistemas mais rígidos como proteínas, não produz maiores interferências. Entretanto, para sistemas muito flexíveis como carboidratos, é produzido um conjunto de dados complexo onde as conformações encontradas podem tanto representar uma população real de confôrmeros como descrever algumas conformações médias durante o tempo de medida que não correspondem a conformações existentes em solução (Cumming & Carver, 1987; Woods, 1998; Wormald *et al.*, 2002). Porém, mesmo diante de algumas dificuldades, esta é a técnica preferencial para a determinação de estruturas de glicoproteínas por determinar um conjunto de estruturas com diferentes confôrmeros, que é o que melhor representa os carboidratos (Woods, 1998;).

Desta forma, o estudo da estrutura e da conformação de carboidratos pode se beneficiar enormemente com o emprego de técnicas complementares, capazes de auxiliar no entendimento da relação entre a dinâmica e a função biológica destas moléculas. No contexto da presente Tese, uma área que vem oferecendo importante contribuições pode ser encontrada nas técnicas de modelagem molecular, como simulações por DM (Dwek, 1996; Woods, 1998; Perez & Mulloy, 2005), como será visto a seguir.

1.5 Caracterização conformacional de biomoléculas através de DM

Historicamente, o uso da DM como um procedimento baseado na computação do movimento de átomos em uma molécula remonta da década de 1950 (Margin & Elliot, 2010). Especificamente para o uso em sistemas biológicos, a DM começou a ser utilizada em 1977 (McCammon *et al.*, 1977), com o estudo de um sistema protéico envolvendo o inibidor de tripsina pancreática bovina (TFPI). Atualmente, é uma ferramenta amplamente difundida, empregada na investigação da estrutura e dinâmica de biomoléculas em geral. O emprego de simulações de DM abrange desde a caracterização da interação de compostos às suas proteínas-alvo, desnaturação e enovelamento protéicos (Ponder & Case, 2003), até o estudo de

exopolissacarídeos bacterianos (Castro *et al*., 2009) e do movimento de tRNAs no ribossomo (Sanbonmatsu *et al*., 2005).

Basicamente, a DM resolve as equações do movimento de Newton em função do tempo:

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

, *s*endo $d^2r_i(t)/dt^2$ a aceleração de uma partícula *i*, m_i sua massa e F_i a força sobre este átomo em um determinado instante *t*. Quando esta equação é resolvida para uma sucessão de instantes *t*, sobre todos os átomos do sistema, gera uma trajetória de movimento das moléculas em estudo, em nível atomístico, como função do tempo (Leach, 2001).

A integração dos componentes é realizada de forma que uma força F_i acarreta uma aceleração sobre um determinado átomo *i* e, em conseqüência, causa uma mudança de sua posição num intervalo de tempo Δt relativo à aceleração. Esta equação, entretanto, não é capaz de determinar a magnitude e a direção da força F_i sobre os átomos do sistema, nem sua relação com as características químicas de cada molécula em estudo (Leach, 2001).

O cálculo destes parâmetros é feito em função de mudanças nas energias potencial e cinética (V) entre a posição atual e a posição seguinte (a que representará o próximo passo da simulação, e assim sucessivamente) sobre cada átomo separadamente, a partir de suas coordenadas atômicas (de r_i a r_n), sendo então descrita por, desta forma a equação anterior pode ser reescrita como:

$$F_i = \frac{-\partial V(r_{i,\ldots} r_n)}{\partial r_i}$$

Esta equação irá representar uma superfície de energia potencial relacionada a cada tipo de molécula, sendo descrita pelo denominado Campo de Força (V) (Schlick, 2006).

O campo de força pode ser definido como um conjunto de funções e parametrizações usadas em cálculos de mecânica molecular (de Sant'Anna, 2002), incluindo, no caso do campo de força GROMOS96 43a1 (van Gunsteren *et al.*, 1996), empregado na presente Tese. Então o campo de força em questão será definido pelos seguintes somatórios:

Termos de estiramento de ligação química (b):

$$\sum_{ligação} 1/2K_b[b-b_0]^2$$

Somando-se a este temos os termos de estiramento dos ângulos de ligação de valência (θ):

$$\sum_{ângulo} 1/2K_\theta[\theta-\theta_0]^2$$

Termos quiralidade ou de definição de deformação de átomos em um plano ou estereoquímica (diedros impróprios) (ξ):

$$\sum_{\substack{diedros\\ improprios}} 1/2K_\xi[\xi-\xi_0]^2$$

E finalmente o termo para torção dos ângulos de diedros próprios (δ):

$$\sum_{diedros} K_\varphi[1+\cos(n\varphi-\delta)]$$

Além destas temos os termos de interações intermoleculares, que são termos não ligados como interações coulômicas e interações de van deer Waals:

$$\sum_{pares(i,j)}[C_{12}(i,j)/r_{ij}^{12}-C_6(i,j)/r_{ij}^6+q_iq_j/4\pi\varepsilon_0\varepsilon_r r_{ij})]$$

A soma destes fatores corresponde ao termos do campo de força (V).

Diversos campos de força estão disponíveis para simulações de DM, tendo sido a maioria parametrizada para o estudo de proteínas. Dentre estes, merecem destaque, por seu amplo uso na literatura, AMBER (Case *et al.*, 2005), CHARMM (MacKerell *et al.*, 1998), CVFF (Kitson & Hagler, 1988), TRIPOS (Clark *et al.*, 1989), OPLS (Jorgensen *et al.*, 1988) e GROMOS96 (van Gunsteren *et al.*, 1996).

No entanto, embora a maioria dos campos de força disponíveis seja capaz de lidar com proteínas, existe uma carência de parâmetros específicos para outras classes de biomoléculas, tais como ácidos nucléicos, lipídios, compostos sintéticos e, especificamente no âmbito da presente Tese, carboidratos (Pérez *et al.*, 1998). Neste caso, a presença de diversos grupos altamente polares, a elevada flexibilidade das ligações glicosídicas e as diferenças nas propriedades eletrônicas que podem ocorrer durante modificações conformacionais dificultam a parametrização de campos de força para estes compostos (Pérez *et al.*, 1998).

Um dos maiores desafios associados à caracterização conformacional de carboidratos, tanto do ponto de vista experimental quanto teórico, envolve a ligação glicosídica, ponto de conexão entre duas unidades monossacarídicas extremamente

flexível (Woods, 1998). Enquanto alguns campos de força colocaram parâmetros adicionais para descrever glicídeos, outros buscaram aumentar sua acurácia se dedicando especificamente para descrever carboidratos e, por conseguinte, limitando sua habilidade em descrever os outros tipos de moléculas, como peptídeos e proteínas em glicopeptídeos e glicoproteínas, respectivamente (Rutherford *et al.*, 1993).

Dentre os campos de força utilizados para determinação da conformação de carboidratos existe um tipo que considera apenas as interações de van der Waals, com exceção apenas ao ângulo. Este modelo, conhecido como campo de força de esfera rígida exo-anomérica (HSEA) (Bock, 1983; Thogersen *et al.*, 1982), obteve modelos para monossacarídeos e dissacarídeos compatíveis com resultados de RMN (Thogersen *et al.*, 1982; Lemieux & Bock, 1983), embora ignore completamente os efeitos eletrostáticos, como as ligações de hidrogênio e interações dipolares (Woods, 1998).

Outra classe de campos de força aplicados a glicanas são os campos de força usualmente aplicados para a simulação de proteínas, onde a energia potencial é expressa como somatório das componentes de ligação, ângulos, diedros e interações não-ligadas. A partir da inclusão de parâmetros específicos para carboidratos, os campos de força AMBER (Case *et al.*, 2005), CHARMm (MacKerell *et al.*, 1998), GROMOS (van Gunsteren *et al.*, 1996), CVFF (Kitson & Hagler, 1988) e TRIPOS (Clark *et al.*, 1989) também foram empregados com sucesso para a modelagem deste tipo de biomoléculas (Woods, 1998). Entretanto de uma maneira geral, estudos comparando alguns dos principais campos de força para carboidratos atualmente disponíveis indicam que nenhum deles é consistentemente melhor que os demais, mas que abordagens específicas, seguidas por cada um deles, podem ser melhores para estudar diferentes aspectos desses compostos (Martín-Pastor *et al.*, 1996; Peréz *et al.*, 1998; Corzana *et al.*, 2003; Hemmingsen *et al.*, 2004).

Considerado uma exceção entre os campos de força para carboidratos, os campos MM2 (Allinger, 1977) e MM3 (Allinger *et al.*, 1989) têm sido utilizados para o estudo das conformações de mono e dissacarídeos. São campos de força com maior complexidade matemática, o que permite uma melhor avaliação do geometria dos carboidratos em relação aos campos de força acima citados. Entretanto, estes campos possuem uma limitação em tratar de simulações em ambiente aquoso com

qualquer modelo de água disponível, tendo seu uso limitado a estudo de fase gasosa e minimização de energia (Woods, 1998).

Neste contexto os mapas de contorno produzidos nos campo de força MM2/3 são referência em se tratando de conformações de dissacarídeos. O portal GLYCOSCIENCES.de (Lutteke *et al.*, 2006) possui um banco de dados com diversos mapas de contorno para dissacarídeos produzidos nestes campos de força, assim como para outros campos de força, chegando a em torno de 3000 mapas, entretanto a grande maioria destes mapas é de dissacarídeos idênticos e geralmente gerados nos campos de força MM2/3.

Nesse sentido, o Grupo de Bioinformática Estrutural vem se dedicando ao uso e desenvolvimento de estratégias de modelagem molecular úteis na descrição e caracterização de carboidratos e glicoconjugados em geral. Esse empenho centra-se no emprego do campo de força GROMOS96 43a1 (van Gunsteren *et al.*, 1996), um campo de força do tipo átomo unido que, por conseguinte, reduz o custo computacional associado aos cálculos, dentro da plataforma GROMACS (van der Spoel *et al.*, 2005). Este pacote de programas é conhecido na literatura como um dos mais rápidos, se não o mais rápido, dedicado à realização de simulações de DM. Esta associação de campo de força e programa ágeis nos permite obter resultados robustos para grandes sistemas moleculares em um tempo reduzido e empregando computadores de baixo custo.

Contudo, embora diversos avanços tenham sido realizados na parametrização de carboidratos para o GROMOS96 (van Gunsteren *et al.*, 1996), alguns substituintes diferentes de hidroxilas ainda não possuem parâmetros no campo de força. Desta forma, o grupo de pesquisas empregou um esquema de cargas alternativo para a descrição conformacional de carboidratos complexos como glicosaminoglicanos (Verli & Guimarães, 2004; Becker *et al.*, 2005), possibilitando a reprodução conformacional de moléculas como a heparina (Verli & Guimarães, 2004). Nestes estudos criaram-se as condições necessárias para a descrição do complexo entre um pentassacarídeo sintético, derivado de heparina, e antitrombina (Verli & Guimarães, 2005), da mesma forma que a realização de estudos envolvendo a identificação das forças responsáveis pelo equilíbrio conformacional do resíduo IdoA, presente na estrutura da heparina (Pol-Fachin & Verli, 2008). Adicionalmente, outros trabalhos nessa linha incluem a caracterização da

conformação de α-fucanas e α-galactanas (Becker *et al.*, 2007) e de uma β-galactana sulfatada (Castro *et al.*, 2009), originados de ouriços-do-mar .

Embora o desenvolvimento e aprimoramento de campos de força para carboidratos estejam crescendo gradativamente nos últimos anos, a adequação desses parâmetros no estudo de glicoconjugados, tais como glicoproteínas, ainda encontra-se bastante incipiente. Isso pode ser evidenciado pelo pequeno número de estudos envolvendo DM de glicoproteínas (Tabela 3), incluindo: (1) uma lectina de *Erythrina corallodendron* (Naidoo *et al.*, 1997) e glicopeptídeos derivados da proteína viral gp120 (Huang *et al.*, 1997), utilizando o campo de força CHARMM (Ha *et al.*, 1988); (2) MHC de classe I (Mandal & Mukhopadhyay, 2001), através dos parâmetros do CVFF (Hwang *et al.*, 1998); e (3) príons de camundongo (Wong *et al.*, 2000) e humano (Zuegg & Gready, 2000), uma glicoproteína anticongelamento (Nguyen *et al.*, 2002), o fator da coagulação humana VII (Perera *et al.*, 2002), glicopeptídeos derivados de hemaglutinina (Bosques *et al.*, 2004) e mucina humana (Rubinsten *et al.*, 2004), utilizando campos de força AMBER (Woods *et al.*, 1995; Kirschner *et al.*, 2008).

Tabela 3: Estudos envolvendo DM de glicoproteínas

Glicoproteína / glicopeptídeo[a]	Glicosilação	Campo de Força
Lectina de *E. corallodendron*	N-	CHARMM
Glicopeptídeos derivados de gp120	N- e O-	CHARMM
MHC de classe I	N-	CVFF
Príon de camundongo	N-	AMBER
Príon humano	N-	AMBER - GLYCAM
Glicoproteína anticongelamento 8	O-	AMBER - GLYCAM
fVII da coagulação humana	O-	AMBER - GLYCAM
Glicopeptídeo derivado de hemaglutinina	N-	AMBER - GLYCAM
Glicopeptídeo derivado de mucina humana	O-	AMBER - GLYCAM
Várias	N- e O-	GROMACS – GROMOS43a1
α_1 acetilcolina nicotínica	N-	AMBER - GLYCAM

[a] Referências, em ordem de aparecimento: Naidoo *et al.*, 1997; Huang *et al.*, 1997; Mandal & Mukhopadhyay, 2001; Wong *et al.*, 2000; Zuegg & Gready, 2000; Nguyen

et al., 2002; Perera *et al.*, 2002; Bosques *et al.*, 2004; Rubinsten *et al.*, 2004; Pol-Fachin, *et al.*, 2009; Dimitropoulos *et al.*, 2011.

Por fim, mesmo que os campos de força sejam capazes de descrever adequadamente a conformação de glicoproteínas, a aplicação das ferramentas de modelagem molecular de forma mais ampla no estudo da relação entre a estrutura e conformação de glicoproteínas as suas respectivas funções biológicas envolve um problema adicional: como as estruturas tridimensionais destas moléculas serão obtidas? Para a parte protéica, as fontes de informação são diversas, incluindo-se a cristalografia de raios-X, RMN ou mesmo modelagem comparativa. Contudo, as principais fontes de informação estrutural para carboidratos não possuem resolução atomística. Isto cria uma demanda e, ao mesmo tempo, uma oportunidade: o desenvolvimento de estratégias confiáveis para a obtenção de modelos tridimensionais de alta qualidade para carboidratos, capazes de representarem seus estados conformacionais majoritários em soluções biológicas e servirem como ponto de partida para simulações de DM, pode gerar um importante impacto no conhecimento da biologia estrutural de glicoproteínas.

2 Objetivos

A partir do exposto, a presente Tese tem como objetivo geral contribuir na obtenção e validação de métodos computacionais capazes de auxiliar na caracterização da biologia estrutural de carboidratos, glicoconjugados e, principalmente, glicoproteínas. Para tanto, as seguintes metas foram identificadas:

- Caracterização da acurácia dos parâmetros empregados pelo Grupo de Bioinformática Estrutural na reprodução da conformação de glicoproteínas previamente descritas por RMN;
- Avaliar a possibilidade do emprego do perfil conformacional de dissacarídeos isolados em solução como referência geométrica para a construção de glicanas complexas ligadas à proteínas. Esta hipótese parte do pressuposto de que o arcabouço protéico não provocaria maiores mudanças conformacionais na estrutura sacarídica, hipótese esta que será verificada;
- Aplicar esta metodologia no estudo da AGP, escolhida como caso de estudo. Esta parte da Tese inclui, ainda, as seguintes etapas:
    - Obtenção de um modelo completo para a AGP, em nível atomístico, incluindo diferentes glicoformas;
    - Caracterização, em nível atômico, do complexo formado pela AGP e a P-selectina, proposto como potencialmente envolvido em eventos imunomodulatórios.

3 Metodologia

3.1 Programas utilizados

Diversas metodologias de modelagem molecular foram utilizadas no presente trabalho, incluindo MM, DM e estudos de *docking*. Os protocolos referentes a cada um destes métodos estão descritos em detalhes a seguir. Mas, de forma global, os programas utilizados incluem:

- Ferramentas de visualização de moléculas: VMD v1.8.6 (Humphrey *et al.*, 1996), PyMol v1.3 e MOLDEN (Schaftenaar & Noordik, 2000);
- Construção de mapas de contorno e simulações de DM: pacote GROMACS v3.3.3 (van der Spoel *et al.*, 2005);
- Geração de topologias para carboidratos: PRODRG *Beta* v2.5 (Schuettelkopf & van Aalten, 2004);
- Cálculos de *docking* molecular: AutoDock 3.0 (Moris *et al.*, 1998).

3.2 Sistemas simulados

Dentre os sistemas simulados para os dissacarídeos e para as glicoproteínas foram 25 sistemas (Tabela 4). Para os mínimos de energia foram simulados mais de uma conformação dependendo dos resultados dos mapas de contorno. Para as simulações de glicoproteínas a partir de dados de RMN mais de uma estrutura foi simulada no caso de conformações de glicanas muito diferentes.

3.3 Nomenclatura e definições

Todos os dissacarídeos envolvidos nesta tese utilizaram nomes e simbologia conforme as recomendações da IUPAC (IUPAC, 1996). Tanto a numeração dos resíduos como a definição da ligação entre duas unidades de monossacarídeos (ponte glicosídica) seguiram as mesmas normas.

A orientação relativa de um par de monossacarídeos ligados foi definida por dois ou três diedros, onde uma ligação do tipo (1→X) onde X pode ser 2, 3, 4 ou 6 é definida para o par de diedros ϕ e ψ torsionais desta ligação como:

$\phi = O_5\ C_1\ O_{x'}\ C_{x'}$

$\psi = C_1\ O_{x'}\ C_{x'}\ C_{x'-1}$

Tabela 4: Dissacarídeos e glicoproteínas estudados

Sistema	Estruturas simuladas	Tempo total de cada simulação
α-D-Man-(1→2)-α-D-Man	2	0.1 μs
β-D-GlcNAc-(1→2)-α-D-Man	2	0.1 μs
α-L-Fuc-(1→3)-β-D-GlcNAc	2	0.1 μs
α-D-Man-(1→3)-α-D-Man	2	0.1 μs
β-D-Gal-(1→4)-β-D-GlcNAc	2	0.1 μs
β-D-Man-(1→4)-β-D-GlcNAc	2	0.1 μs
β-D-GlcNAc-(1→4)-β-D-GlcNAc	4	0.1 μs
β-D-GlcNAc-(1→4)-β-α-D-Man	1	0.1 μs
α-L-Fuc-(1→6)-β-D-GlcNAc	3	0.1 μs
α-D-Man-(1→6)-α-D-Man	2	0.1 μs
α-D-NeuAc-(2→3)-β-D-Gal	3	0.1 μs
Domínio EGF-like não glicosilado – PDB 1F7E	1	0.05 μs
CD59 não glicosilado – PDB 1CDQ	1	0.05 μs
Domínio de adesão CD2 não glicosilado – PDB 1CDB	1	0.05 μs
α-hCG não glicosilada – PDB 1DZ7	1	0.05 μs
Domínio EGF-like glicosilado – PDB 1FF7	1	0.05 μs
CD59 glicosilada – PDB 1CDR	2	0.1 μs
CD59 glicosilada – PDB 1CDS	1	0.05 μs
Domínio de adesão CD2 glicosilado – PDB 1GYA	2	0.1 μs
α-hCG glicosilada – PDB 1HD4	2	0.1 μs
COX-1 glicosilada	1	0.05 μs
COX-2 glicosilada	1	0.05 μs
AGP não glicosilada – PDB 3BX6	1	0.1 μs
AGP glicosilação com fucoses	1	0.1 μs
AGP glicosilação sem fucoses	1	0.1 μs
AGP – P-selectina – ligação *via* Asn54	1	0.05 μs
AGP – P-selectina – ligação *via* Asn75	1	0.05 μs

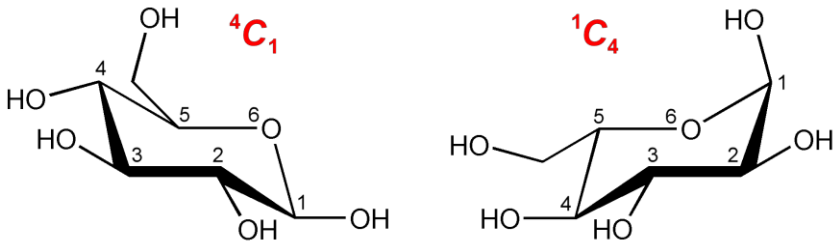
Para ligações do tipo (1→6) há ainda uma definição adicional do ângulo ϖ, devido a ponte glicosídica desta ligação ser mais longa que as demais:

$\varpi = O_{6'}\ C_{6'}\ C_{5'}\ C_{4'}$

3.4 Construção de topologias

As topologias das glicanas utilizadas neste trabalho foram obtidas a partir da ferramenta de construção de topologias PRODRG (Schuettelkopf & van Aalten, 2004). O PRODRG é um servidor que funciona na rede mundial de computadores, gerando um arquivo de topologia para os sistemas de interesse a partir de um arquivo PDB que contenha sua estrutura. Alem dos parâmetros obtidos no servidor as cargas atômicas foram substituídas por cargas calculadas por métodos *ab initio* (Verli & Guimarães, 2004; Becker *et al.*, 2005), e também a adição de diedros impróprios (Tabela 6), necessários para preservar os estados conformacionais ($^4C_1$ ou $^1C_4$) dos monossacarídeos. Uma tentativa de simulação sem os diedros impróprios revelou uma distorção em vários anéis de monossacarídeos, sendo então decidido manter os mesmos. As topologias de cada monossacarídeo empregado, estão disponíveis no final desta Tese (anexos 8.2 e 8.3).

Tabela 5: Definição dos diedros impróprios utilizados para definir a conformação dos monossacarídeos contidos nos sistemas simulados.



Seqüência de átomos definindo o diedro[a]	Ângulos (em graus)	
	$^4C_1$	$^1C_4$
5 – 2 – 4 – 1	2,0	-2,0
5 – 2 – 3 – 1	23,0	-23,0
5 – 2 – 3 – 6	-2,0	2,0

[a] Na figura, a glicose é utilizada como modelo.

[b]Adaptado de Verli, 2005.

3.5 Construção de mapas de contorno relaxado

A partir da construção das topologias de cada monossacarídeo, as preferências conformacionais de dissacarídeos foram avaliadas através de mapas de contorno, análogos aos mapas de Ramachandran. Ao todo, durante a presente Tese, foram estudados 11 dissacarídeos, resultando em 13 mapas de contorno (para as ligações 1→6 são gerados dois mapas por dissacarídeo), a saber: (Tabela 4).

Dessa forma, no caso de ligações 1→2, 1→3 e 1→4, os ângulos ϕ e ψ foram rotados de -180º a 150º, em passos de 30º, gerando um total de 144 confôrmeros para cada dissacarídeo. No caso de ligações 1→6, como a ligação glicosídica é composta por três diedros: ϕ, ψ e ϖ, foram gerados 144 confôrmeros para cada par de ângulos, ou seja, 144 para o par ϕ *versus* ψ e 144 para o par ψ *versus* ϖ. Para todos esses tipos de ligação, cada um destes confôrmeros é então minimizado utilizando-se restrições (diedros impróprios) somente para os diedros da ligação glicosídica em estudo. Em outras palavras, o par de diedros sendo avaliado é mantido preso em cada uma das 144 combinações avaliadas, enquanto o restante da molécula encontra-se livre para ter sua energia minimizada. O protocolo utilizado para a minimização de cada confôrmero envolve o emprego do algoritmo de *steep* e, a partir do mínimo de energia assim obtido, uma simulação de DM com duração de 20 ps a temperatura de 10K e tempo de integração de 0,5 fs para reforçar a busca pelo arranjo mais estável da conformação avaliada (Becker *et al.*, 2007; Pol-Fachin & Verli, 2008). Os valores de energia das 144 conformações de cada par de diedros são então empregados na construção de um mapa de contorno, que irá definir a estabilidade relativa à conformação mais estável de todas (Figura 8).

3.6 Construção da estrutura glicosídica

As estruturas glicosídicas utilizadas nas simulações da AGP glicosiladas foram construídas no site *GLYCOSCIENCES.de* (Lutteke *et al.*, 2006) a partir da ferramenta *GlyProt*. A estrutura glicosilada foi montada a partir da submissão à ferramenta da estrutura cristalográfica da AGP (3BX6.pdb). O *GlyProt* identifica os possíveis sítios de N-glicosilação (no caso, cinco sítios). A partir de outra ferramenta podem ser submetidos a sequência de e mediante a submissão à ferramenta da sequência (em formato de texto) a ser acrescida em cada sitio é montada a estrutura glicosilada que é devolvida como um arquivo de formato .pdb.

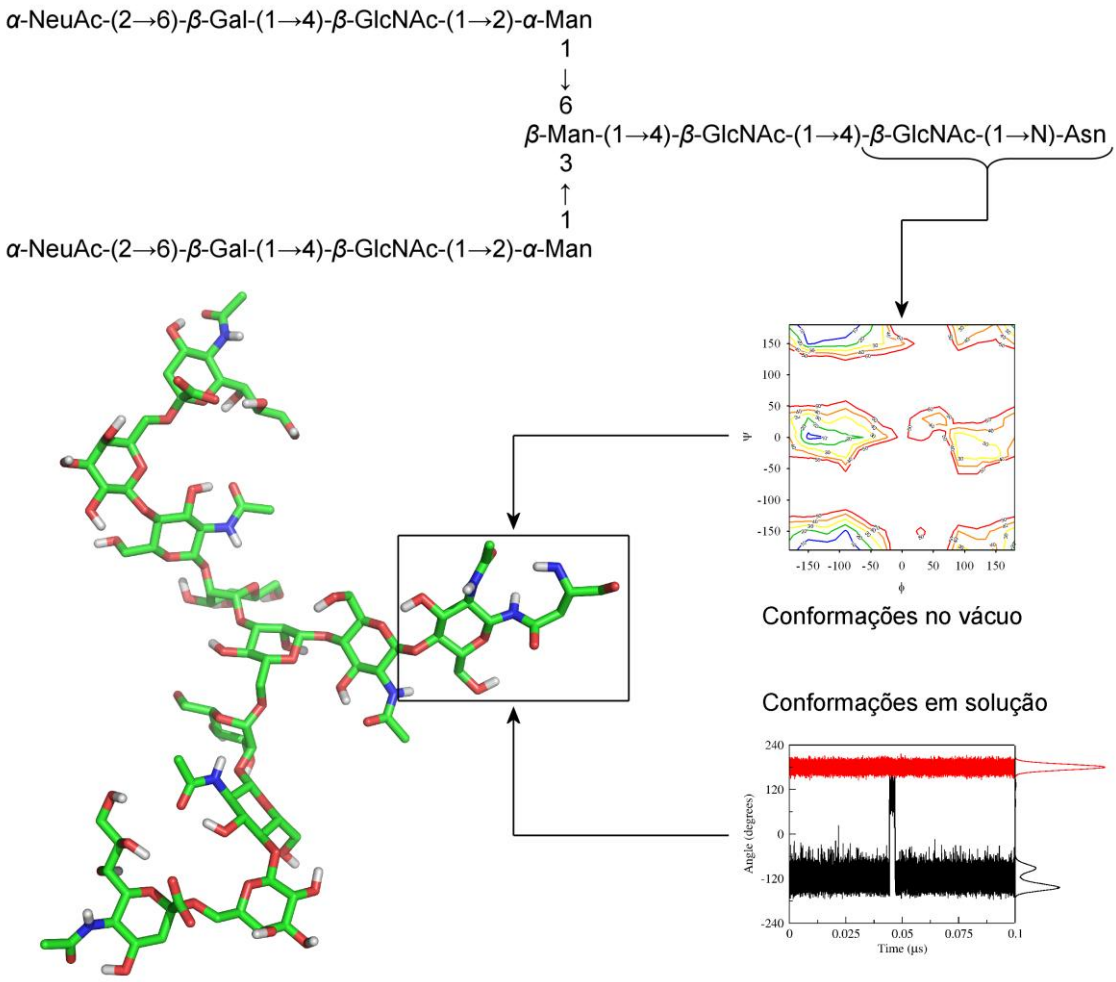


Figura 8: Esquema de como são feitas as simulações dos dissacarídeos. Inicialmente se simula o dissacarídeo no vácuo com as restrições nos ângulos ϕ e ψ, após segue-se outra simulação do mesmo dissacarídeo em solução (Adaptado de Pol-Fachin & Verli, 2011).

3.7 Simulações de DM

3.7.1 Protocolo de simulação

Todas as simulações deste trabalho foram rodadas no pacote GROMACS v. 3.3.3 (van der Spoel *et al.*, 2005) com a campo de força modoficado do GROMOS43a1 .O protocolo geral de simulação foi baseado em procedimentos previamente descritos (de Groot & Grubmüller, 2001). Todas as simulações foram realizadas à temperatura fisiológica (310K), com duração variável, de acordo com o sistema em estudo (de 50 a 100 ns) (Tabela 4). Cada molécula (proteína,

glicoproteína ou dissacarídeo) foi solvatada numa caixa triclínica utilizando condições periódicas de contorno, empregando o modelo de água SPC (Berendsen *et al.*, 1987). Contra-íons (cloreto ou sódio) foram adicionados, conforme a necessidade, de forma a neutralizar as cargas dos sistemas estudados. 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 (PME, Darden *et al.*, 1993), utilizando raios de corte de *Coulomb* e de van der Waals de 9 Å. A temperatura e a pressão do sistema foram mantidas constantes através do acoplamento do soluto, íons e solvente a banhos externos de temperatura e pressão, utilizando constantes de acoplamento de, respectivamente, τ = 0,1 ps e τ = 0,5 ps (Berendsen *et al.*, 1984). A constante dielétrica do meio foi tratada como ε = 1.

Como protocolo para a simulação foi feita uma termalização do sistema com o aquecimento gradativo do sistema, visando uniformizar as energias contidas na estrutura cristalográfica ou de RMN e, desta forma, evitar deformações nas moléculas em estudo. Nesta etapa, após 1 ps de restrição de posição a 50K, cada sistema foi aquecido lentamente de 50 K a 300 K, de maneira que, no primeiro passo de 9 ps foi aquecido um incremento de 50 K, posteriormente, em cada um dos cinco passos de 5 ps, há o aumento da temperatura em 50 K (Figura 9). Após a termalização do sistema, a simulação prossegue na temperatura de equilíbrio de 310 K, pelo tempo estipulado.

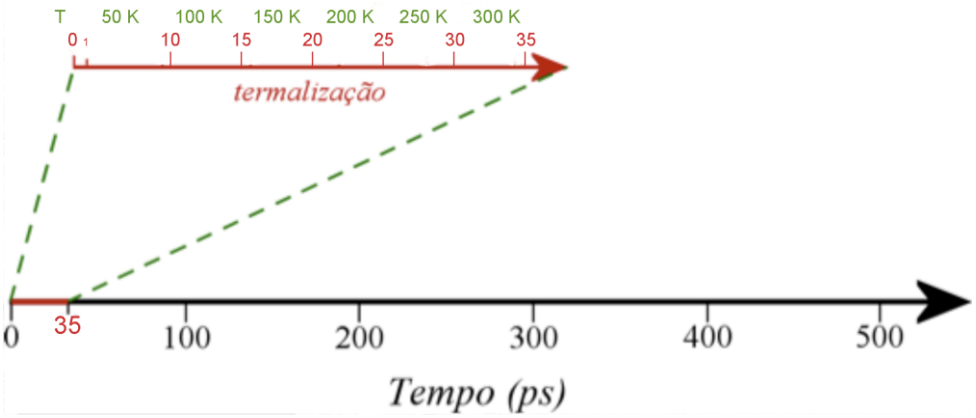


Figura 9: Esquema das etapas que compõe as simulações de DM a 310K, com ênfase na termalização. De 0 ps a 35 ps, uma série de sete etapas, a primeira de 1 ps, a segunda de 9 ps e, após, cinco etapas de 5 ps cada. No procedimento, aquece-se lentamente o sistema em incrementos de 50 K. Após 35 ps, o sistema

está a 300 K e será levado a 310 K no início da DM, se mantendo nesta temperatura pelo resto da simulação (Adaptado de Verli, 2005).

3.7.2 Validação das simulações de DM

A confiabilidade das predições baseadas em cálculos de DM depende, principalmente, da precisão do campo de força utilizado e do tamanho da amostragem do espaço conformacional realizada durante a simulação, que necessita ser suficientemente ampla para descrever as propriedades do sistema em estudo (van Gunsteren & Mark, 1998). De modo geral, o teste final do modelo teórico produzido reside na sua comparação com propriedades experimentais, ou seja, a precisão do método é julgada a partir da observação ou constatação, nas simulações realizadas, de quão precisas são reproduzidas quantidades conhecidas (van Gunsteren & Berendsen, 1990; Karplus & Petsko, 1990). Dessa forma, a validação dos resultados obtidos nas simulações de DM foi realizada através da comparação com dados experimentais prévios, com destaque para estruturas cristalográficas, dados bioquímicos e, principalmente, conformações oriundas de experimentos de RMN, assim como através da observação da estabilidade dos sistemas estudados, ou seja, pela manutenção, durante as simulações, de propriedades que incluem estrutura secundária, energia, densidade e volume.

3.8 Validação do uso dados conformacionais de dissacarídeos na construção de glicanas complexas

A partir dos dados obtidos nas simulações de dissacarídeos em solução onde os confôrmeros mais estáveis foram identificados e dos dados de simulação e RMN de glicoproteínas que continham estes dissacarídeos observamos que é possível reproduzir a conformação dos dissacarídeos componentes de glicoproteínas em simulações com apenas estes dissacarídeos em solução (Fernandes *et al.*, 2010). Desta forma na construção da glicosilação da AGP utilizamos as conformações majoritárias de cada dissacarídeo simulado como conformação inicial das glicanas componentes da AGP. Para tanto foi montado um protocolo onde a glicoproteína foi minimizada mantendo a proteína sob restrições em todos os átomos (constante de força 1000 $kJmol^{-1}$) e procedendo um relaxamento na estrutura glicosídica.

Inicialmente foram acrescentados a topologia os diedros impróprios com os valores de ϕ e ψ de acordo com os resultados das conformações majoritárias dos dissacarídeos com uma constante de força de 334,8 $kJmol^{-1}$ de restrição para estes diedros. Foi feita uma minimização de energia no vácuo com gradiente *steep* de 10000 passos nestas condições. Após outro passo de relaxamento foi submetido ao sistema onde a constante de força foi diminuída a 84,7 $kJmol^{-1}$ rodando outra minimização nas mesmas condições anteriores. Após as restrições aos diedros ϕ e ψ são retiradas e a simulação segue o mesmo roteiro das simulações anteriores (item 3.7).

3.9 Cálculos de docking molecular

Os cálculos de *docking* foram utilizados para a busca de coordenadas passíveis de formar um complexo entre as moléculas de AGP e P-selectina a partir da sequência de glicosilação de SLeX. Partindo da estrutura cristalográfica 1G1S da P-selectina com SLeX foram montadas as conformações iniciais onde a AGP, e mais especificamente o SLeX contido na AGP, mantém uma conformação semelhante a encontrada no cristal da P-selectina. A AGP foi tratada como um ligante rígido para os cálculos de *docking*, que foram feitos apenas para os glicídeos ligados aos resíduos Asn54 e Asn75.

O AutoGrid foi utilizado para a geração dos mapas de *grid* de 0,375 Å com 78x78x92 pontos de *grid* para a molécula de P-selectina. Foi utilizado o algoritmo de *simulated annealing* com os parâmetros padrão seguidos de cem rodadas de *docking*. Os resultados são expressos em uma função de hierarquização que classifica os resultados obtidos de acordo com a melhor energia de interação intermolecular e posições similares.

4 Resultados

4.1 Preâmbulo

Os resultados obtidos serão apresentados a seguir na forma dos trabalhos publicados e submetido durante a realização da presente Tese. Estes trabalhos estão citados abaixo, assim como uma breve descrição sobre cada um.

- Laércio Pol-Fachin, **Cláudia Lemelle Fernandes**, Hugo Verli: GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydr. Res.*, **2009**, *344*; 491-500.

Este artigo avalia a adequação do campo de força GROMOS96 43a1, adicionado por cargas atômicas de Löwdin, na descrição conformacional de glicoproteínas utilizando simulações de DM, tendo como base a comparação com dados experimentais de RMN.

- **Cláudia Lemelle Fernandes**, Liana Guimarães Sachett, Laércio Pol-Fachin, Hugo Verli: GROMOS96 43a1 performance in predicting oligosaccharides conformational ensemble within glycoproteins. *Carbohydr. Res.*, **2010**, *345*; 663-671.

Este trabalho analisa a adequação do uso de conformações obtidas por simulações de DM para dissacarídeos em solução aquosa como pontos de partida para a construção da estrutura de glicanas componentes de glicoproteínas. Estes resultados são comparados com dados de RMN e DM previamente descritos e, por fim, é acrescentado um caso de estudo onde a glicosilação é construída a partir da técnica preconizada.

- **Cláudia Lemelle Fernandes**, Hugo Verli: Structural glycobiology of human α_1-acid glycoprotein and its implication for drug and P-selectin binding. Em preparação.

Este trabalho compõe um caso de estudo de uma glicoproteína altamente glicosilada, onde é analisado o efeito tanto da presença da glicosilação quanto da mudança no padrão de glicosilação sobre a dinâmica da proteína. E adicionalmente é estudado um caso de interação da AGP glicosilada com a P-selectina, esta

interação é relaciona a eventos imunomodulatórios que existem entre a P-selectina e a epítopo específico da glicosilação da AGP.

Este manuscrito está sendo preparado para a submissão à revista *Biochemistry.*

4.2 Trabalho I

Considerando os trabalhos anteriores do grupo no estudo da conformação de glicosaminoglicanos e polissacarídeos sulfatados, o presente trabalho tem por objetivo avaliar o emprego desta metodologia, baseada no campo de força GROMOS96 43a1, adicionado por cargas atômicas de Löwdin na base HF/6-31G**, em reproduzir o perfil conformacional de carboidratos ligados a proteínas previamente estudadas por RMN validando, assim, o emprego destes parâmetros no estudo de glicoproteínas.

Os resultados obtidos indicam que a metodologia empregada é capaz de reproduzir diferentes aspectos conformacionais de glicoproteínas, de acordo com dados experimentais prévios: (1) a conformação de N- e O-ligações glicosídicas, formadas entre aminoácidos e monossacarídeos; (2) a conformação da parte sacarídica de glicoproteínas, de maneira que mais de 95% das geometrias determinadas por RMN para as ligações glicosídicas que compõe tais oligossacarídeos mostraram-se de acordo com os dados de DM; (3) os efeitos da glicosilação sobre a estrutura e conformação da parte protéica de glicoproteínas, conferindo estabilidade conformacional e influenciando propriedades relacionadas a função dessas moléculas.

GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations

Laércio Pol-Fachin, **Cláudia Lemelle Fernandes**, Hugo Verli

Carbohydrate Research, **2009**, *344*; 491-500

Carbohydrate Research 344 (2009) 491–500

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations

Laercio Pol-Fachin [a], Claudia Lemelle Fernandes [a], Hugo Verli [a,b,*]

[a] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil
[b] Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil

ARTICLE INFO

Article history:
Received 3 September 2008
Received in revised form 3 December 2008
Accepted 26 December 2008
Available online 3 January 2009

Keywords:
Glycoproteins
Molecular dynamics
Carbohydrates
Glycosidic linkage
GROMOS96

ABSTRACT

Considering the small number of papers assessing the conformational profile of glycoproteins through molecular dynamics (MD) simulations, the current work reports on a systematic analysis of the performance of the GROMOS96 43a1 force field and Löwdin HF/6-31G**-derived atomic charges in the conformational description of glycoproteins. The results substantiate the accuracy of the computational representation of glycoprotein conformational ensembles in aqueous solution based on their agreement to available experimental information, supporting further contributions of computational techniques, mainly MD, in future studies on the characterization of glycoprotein structure and function.

1. Introduction

Glycoproteins and glycopeptides are known to play important roles in many biological events, such as cell adhesion, cell–cell communication, immune response, intracellular targeting, and protease resistance.[1,2] These glycans can be located at various positions of the protein surface, depending on specific consensus sequences, and are linked only through specific amino acid residues. The linkages most frequently found in nature are the N-glycosidic linkages, involving mainly asparagine (Asn) and *N*-acetyl-D-glucosamine (GlcNAc) residues, and the O-glycosidic linkages, mostly including serine (Ser) or threonine (Thr) and, for instance, D-galactose (Gal), *N*-acetyl-D-galactosamine (GalAc), L-fucose (Fuc), or D-mannose (Man) residues. Structurally, these carbohydrate moieties impact several physicochemical properties of proteins, including hydration and polarity, frequently participating in its folding and conformational stabilization.[1,2]

The comprehension of the three-dimensional structure and the dynamical properties of both protein and carbohydrate moieties of glycoproteins is a requirement for a better understanding of the molecular basis of their interaction with each other, with the surrounding environment, and with their molecular targets. In this context, several experimental methods have been applied to study glycosylated proteins and glycans, such as X-ray crystallography and NMR spectroscopy. Concerning the former technique, the conformational flexibility of the glycan at the protein surface usually prevents such molecules from being crystallized,[3,4] but when crystals are available, the electron density is affected by the high thermal motion of the glycan moiety, compromising the accuracy of the geometry beyond the rigid core region of N-glycans.[5] NMR methods provide the solution ensemble of conformations in a set of average three-dimensional models, but an unambiguous determination of the complete conformational space of glycoproteins is hampered because only from one to three contacts across a given glycosidic linkage are usually detected.[6]

Considering the difficulties associated with the experimental determination of glycoproteins structure and conformation, MD simulations have been used to clarify the dynamic aspects of macromolecules. In particular, such simulations support the study of these molecules mimicking their natural environment and describe their conformational properties with a reasonable level of accuracy. On the other hand, only a small number of studies have assessed the profile of glycoproteins and glycopeptides through MD simulations, employing different set of parameters for their glycan parts, such as AMBER, CHARMm, and CVFF.[7–9] In comparison to such parameters, our group has been working on MD conformational representation of carbohydrates, polysaccharides, and glycosaminoglycans[10–13] based on the GROMOS96 43a1 force field,[14] enhanced by Löwdin HF/6-31G**-derived atomic charges.[10,12] This approach supports the adequate description of such molecules in aqueous solutions.

Therefore, the current work intends to evaluate the capability of the GROMOS96 43a1 force field and Löwdin HF/6-31G**-derived

* Corresponding author. Tel.: +55 51 3308 7770; fax: +55 51 3308 7309.
E-mail address: hverli@cbiot.ufrgs.br (H. Verli).

0008-6215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.12.025

atomic charges, and to represent adequately the conformational ensemble of glycoproteins through MD simulations in aqueous solutions. The results of these simulations were compared to experimental data detailing the geometry of the glycosidic linkages between two monosaccharides and between a monosaccharide and an amino acid. Additionally, the effects of glycosylation upon the protein moiety of the glycoproteins were also assessed. Based on the data, the employed methodology may be expected to further contribute to structural and functional studies of glycoconjugates through reliable and accessible methodology.

2. Experimental

2.1. Computational methods

2.1.1. Nomenclature and software

The nomenclature recommendations and symbols as proposed by IUPAC[15] were used. The relative orientation of a pair of contiguous carbohydrate residues is described, for different types of linkages, by two or three torsional angles at the glycosidic linkage. For a (1→X) linkage, where 'X' is '2', '3', '4' or '6' for the (1→2), (1→3), (1→4) or (1→6) linkages, respectively, the ϕ and ψ are defined as shown below:

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

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

For a (1→6) linkage, the ω is defined as shown below:

$$\omega = \mathrm{O6{-}C6{-}C5{-}C4} \quad (3)$$

The manipulation of structures was performed with MOLDEN[16] and VMD,[17] the secondary structure content analyses were performed with PROCHECK,[18] and all the MD calculations and remaining analyses were performed using GROMACS simulation suite[19] and GROMOS96 force field,[14] with all saccharide topologies generated with PRODRG.[20]

2.1.2. Topology construction

The oligosaccharide fragments present in the glycoproteins under the PDB codes 1CDR,[21] 1CDS,[21] 1GYA,[22] 1HD4,[23] and 1FF7[24] (Fig. 1) were described in GROMOS96 43a1 force field through PRODRG server-derived topologies,[20] which were further refined by atomic charges, as described previously.[10,12] Briefly, each monosaccharide was submitted to HF/6-31G** energy minimization, followed by Hessian matrix analyses to characterize them unequivocally as true minima at the potential energy surface. The residues were divided in charge groups, while improper dihedrals were included as necessary to preserve the hexopyranose conformation of each monosaccharide in accordance with its expected form in aqueous solution: $^4C_1$ for D-GlcNAc, $1C_4$ for L-Fuc, $4C_1$ for D-Man, and $^4C_1$ for D-Gal. Additionally, proper dihedrals, as described in GROMOS96 43a1 force field for glucose, were also included in the PRODRG obtained topologies in order to support stable simulations. Concerning the N- and O-linkages between the carbohydrate and amino acid residues, the atomic charges were calculated for the Asn-carbohydrate (GlcNAc $C_1 = 0.110$, Asn $N_\delta = -0.265$, and Asn $H_{N\delta} = 0.155$) or Ser-carbohydrate (Fuc $C_1 = 0.150$, Ser $O_\gamma = -0.320$,

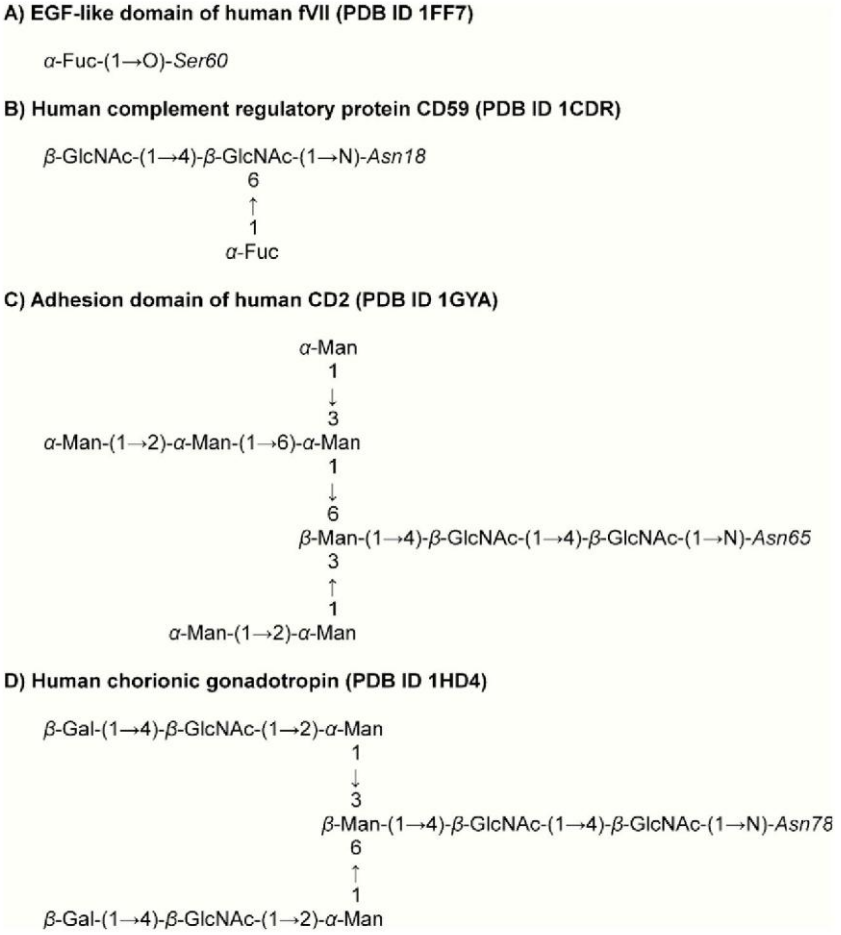


Figure 1. Schematics of the studied glycans and the corresponding glycosylated amino acids in the studied proteins.

and Ser C_β = 0.170) connection charge groups. Such an approach, based on parameters derived from minimum energy conformations, has been successfully applied in previous work in carbohydrate modeling.[10–13,25,26]

2.1.3. MD simulations

The structures of glycoproteins, as obtained from NMR data, were retrieved from the PDB, including the human complement regulatory protein CD59 (CD59), in its non-glycosylated (PDB ID 1CDQ) and glycosylated (PDB codes 1CDR and 1CDS) forms,[21] the α-subunit of the human chorionic gonadotropin (α-hCG), in both non-glycosylated (PDB ID 1DZ7)[27] and glycosylated (PDB ID 1HD4) structures,[23] the first epidermal growth factor-like (EGF-like) domain of the human blood coagulation fVII in its non-glycosylated (PDB ID 1F7E) and glycosylated (PDB ID 1FF7) forms,[24] and the adhesion domain of human CD2 ($hsCD2_{105}$) in its glycosylated (PDB ID 1GYA)[22] and non-glycosylated (PDB ID 1CDB)[28] forms, in a total of five glycosylated (Fig. 1) and four non-glycosylated proteins. These structures were solvated in a rectangular box using periodic boundary conditions and SPC water model.[29] Amino acids ionization was automatically adjusted to the plasma pH, with histidines being non-ionized and at the N_τ–H tautomeric form. Counterions (Na^+ and Cl^-) were added to neutralize the systems, whenever needed. The employed MD protocol was based on previous studies, as described.[10,12] The Lincs method[30] was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using Steepest Descent algorithm. Electrostatic interactions were calculated with Particle Mesh Ewald method.[31] Temperature and pressure were kept constant by coupling glycoproteins, ions, and solvent to external temperature and pressure baths with coupling constants of τ = 0.1 and 0.5 ps,[32] respectively. The dielectric constant was treated as ε = 1, and the reference temperature was adjusted to 310 K. The systems were slowly heated from 50 to 310 K, in steps of 5 ps, each one increasing the reference temperature by 50 K. Each simulation was extended to 50 ns, without any restraint, while a reference value of 3.5 Å between heavy atoms was considered for a hydrogen bond and a cutoff angle of 30° was used between hydrogen–donor–acceptor.[19] As the NMR selected structures include several models in the same PDB file, a total of eight glycoproteins (one model for 1CDS and 1FF7, and two models for 1CDR, 1GYA, and 1HD4), plus four non-glycosylated counter-parts were simulated to achieve a broader conformational description through MD. Details of the simulated models are included in Table 1. Additional simulations of the non-glycosylated proteins were performed with OPLS-AA force field in order to check the pattern of secondary structure as observed in GROMOS96 MD.

3. Results and discussion

3.1. Simulation systems

The analysis of GROMOS96 43a1 force field performance on glycoproteins conformational description was based on the MD simulation of four different proteins: (1) the EGF-like domain of human fVII; (2) the human complement regulatory protein CD59; (3) the adhesion domain of human CD2; and (4) the α-subunit of the human chorionic gonadotrophin. These proteins were simulated in both glycosylated and non-glycosylated forms, taking their NMR-derived model as the starting geometry. Among these molecules, three have N-linked glycans and one has an O-linked glycan. These systems were analyzed taking the NMR-obtained ensemble from each macromolecule as reference, to supply the information of GROMOS96 43a1 force field performance on the description of glycan–protein linkage of different glycosidic linkages composing the glycan structures, and the influence of glycosylation on the protein moiety of each molecule.

Actually, conformational data derived from single crystals of glycoconjugates should be considered carefully since the crystal environment may supply distorted hexopyranose rings and/or be submitted to packing effects, which are also known to occur for other classes of molecules, such as DNA[33] and proteins.[34,35] In contrast, although NMR techniques are not considered to provide an unambiguous determination of the complete conformational space of glycans,[3] it does provide solution conformations for the studied glycans which, in a set of average three-dimensional structures, could provide a realistic sample of glycans-composing disaccharide conformations and, consequently, a better reference for the experimental support of MD in solutions.

3.2. N- and O-glycosidic linkage geometry

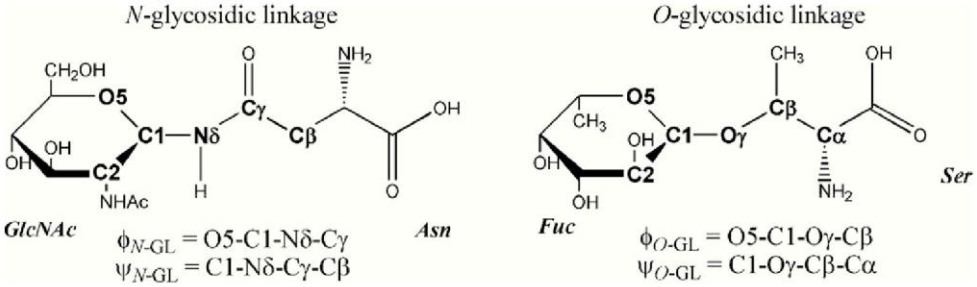
One of the main features determining the orientation of carbohydrates attached to proteins lies in its linkage to specific amino acid residues, such as asparagine and serine/threonine. The N-glycosidic linkages have been extensively studied by both NMR and X-ray methods, while a lower amount of experimental data is available for the O-glycosidic linkages, probably due to the increased flexibility and versatility of these linkages, involving distinct amino acid and monosaccharide residues. Therefore, the performed simulations of glycoproteins were initially analyzed by considering the connection between carbohydrates and proteins, using a set of experimental data as reference, as shown in Table 1 for average geometries and in Figure 2 for the respective conformational behavior of each linkage.

As can be observed, previous analysis of X-ray data indicates that the $\phi_{N\text{-GL}}$ dihedral angle presents higher flexibility compared to the $\psi_{N\text{-GL}}$ dihedral angle, with most X-ray structures presenting the $\phi_{N\text{-GL}}$ with a 80° amplitude (between −140° and −60°) and $\psi_{N\text{-GL}}$ with a 40° amplitude (between 160° and −160°).[36] Additional data, based on ~500 crystal structures obtained from the PDB, indicate a lower flexibility, in which the GlcNAc-(1→N)-Asn linkage exhibits an average $\phi_{N\text{-GL}}$ geometry at ~−100° and $\psi_{N\text{-GL}}$ at ~180°.[37] One of the most recent papers upon such linkages, combining X-ray crystallography and both HF and B3LYP with the 6-31+G* basis set, on models and analogs of GlcNAc-(1→N)-Asn,[38] also observed similar geometries for such dihedral angles. This large amplitude of the $\phi_{N\text{-GL}}$ also suggests the occurrence of multiple conformer populations co-existing in solution and being sustained depending on the crystallographic environment, as indeed observed in MD (Fig. 2A).

Hence, all glycoproteins presented equivalent conformational patterns around the GlcNAc-(1→N)-Asn linkage during the performed simulations, in a behavior apparently independent on the surrounding protein scaffold (Fig. 2A and Supplementary data). The curve representing such conformational patterns include about 81% of the NMR experimental geometries of $\phi_{N\text{-GL}}$, as well as 99% in $\psi_{N\text{-GL}}$ (in a total of 226 geometries for each dihedral angle),[21–23,39–42] indicating a good reliability of the simulations to predict the dynamics of N-linked glycans.

The geometry of O-glycosidic linkages was also evaluated and compared to three NMR studies, containing a fucose residue linked to two amino acid residues, Ser[24] or Thr,[43] as well as an additional structure presenting a 3-*O*-methyl-mannose residue linked to Thr.[44] It may be observed that the $\phi_{O\text{-GL}}$ and $\psi_{O\text{-GL}}$ dihedral angles of these linkages adopt similar geometries regardless of whether Ser and Thr is the involved amino acid residue (Table 1). Different from N-glycosidic linkages, no major flexibility differences could be observed when comparing $\phi_{O\text{-GL}}$ and $\psi_{O\text{-GL}}$ dihedral angles. In addition, the $\phi_{O\text{-GL}}$ presented two major MD conformations, which are both found in NMR data, around −70° (in 1URK and 1FF7 PDB

Table 1
Comparison of the dihedral angles around N- and O-glycosidic linkages, including X-ray, NMR, and MD simulations data



Glycosidic linkage	Structures	Dihedral angle (°)[†]			
		$\varphi_{N\text{-GL}}$	$\psi_{N\text{-GL}}$	$\varphi_{O\text{-GL}}$	$\psi_{O\text{-GL}}$
GlcNAc-(1→N)-Asn	X-ray[a]	−60 to −140	160 to −160	—	—
	X-ray[b]	−99 ± 21	177 ± 12	—	—
	NMR[c]	−91 ± 45	180 ± 7	—	—
	NMR—1CDR[d]	−77 ± 45	180 ± 0	—	—
	MD—1CDR[e,*]	−121 ± 29	−179 ± 14	—	—
	MD—1CDR[f,*]	−126 ± 28	179 ± 13	—	—
	NMR—1CDS[d]	−63 ± 44	180 ± 0	—	—
	MD—1CDS[g,*]	−119 ± 30	−177 ± 14	—	—
	NMR—1GYA[d]	−99 ± 15	179 ± 2	—	—
	MD—1GYA[h,*]	−121 ± 28	178 ± 13	—	—
	MD—1GYA[i,*]	−116 ± 29	−174 ± 14	—	—
	NMR—1HD4[d]	−86 ± 50	177 ± 6	—	—
	MD—1HD4[j,*]	−120 ± 30	−174 ± 14	—	—
	MD—1HD4[k,*]	−116 ± 30	−178 ± 12	—	—
Fuc-(1→O)-Ser	NMR—1FF7[d]	—	—	−67 ± 10	178 ± 10
	MD—1FF7[l,*]	—	—	−110 ± 29	178 ± 22
Fuc-(1→O)-Thr	NMR—1URK[d]	—	—	−104 ± 30	−154 ± 76
3-*O*-methyl-Man-(1→O)-Thr	NMR—2HGO[d]	—	—	72 ± 3	120 ± 22

[†] Values presented as averages ± standard deviation values, obtained from all trajectory points, that is, no distinction of conformer populations was made to support a direct comparison to experimental data.
[*] 50 ns MD averages.
[a] Data from Ref. 36.
[b] Data from Ref. 37, wherein the geometry from most populous conformer is present.
[c] Average dihedral angle values from all models included in PDB codes: 1CDR,[21] 1CDS,[21] 1GYA,[22] 1HD4,[23] 1BZB,[39] 1BYV,[39] 2FN2,[40] 1E88,[41] 1E9J,[23] 2JXA,[42] and 2JX9.[42]
[d] Average dihedral angle values from all models included in PDB codes 1CDR and 1CDS (10), 1URK (15), 1GYA (18), 1FF7 and 2HGO (20) and 1HD4 (26).
[e] Using the model 1 from PDB file 1CDR as a starting structure.
[f] Using the model 8 from PDB file 1CDR as a starting structure.
[g] Using the model 1 from PDB file 1CDS as a starting structure.
[h] Using the model 1 from PDB file 1GYA as a starting structure.
[i] Using the model 12 from PDB file 1GYA as a starting structure.
[j] Using the model 1 from PDB file 1HD4 as a starting structure.
[k] Using the model 11 from PDB file 1HD4 as a starting structure.
[l] Using the model 1 from PDB file 1FF7 as a starting structure. Values of 111 ± 28° were also populated in the simulation.

files) or around 70° (in 2HGO PDB file). Such a pattern is also in good agreement with the distribution of the dihedral angles around the Fuc-(1→O)-Ser linkage of the EGF-like domain, where two main peaks are indeed observed around the $\phi_{O\text{-GL}}$ angle (Fig. 2B). Again, good reliability is observed for O-linked glycoproteins, with 97% of the experimental geometries within the distribution curve for $\phi_{O\text{-GL}}$ and 69% for $\psi_{O\text{-GL}}$ (in a total of 45 geometries for each dihedral angle).[24,43,44]

3.3. Dynamics of glycosidic linkages that compose the glycan part of glycoproteins

Similar to the conformational behavior observed for the linkage between monosaccharides and amino acids, a given disaccharide showed mostly equivalent conformational profiles, with a minor influence of the protein scaffold on glycan conformational preference (Fig. 3 and Supplementary data). As observed, most of the experimental conformations, defined by NMR data, are within the distribution curves obtained from the MD simulations, indicating an adequate reproduction of the glycan conformational ensemble. In addition, considering that NMR methods provide average three-dimensional models, giving rise to ensemble-average properties instead of single solution conformations,[6] it is feasible to observe NMR geometries between main conformational states/peaks. Such characteristics also correlate to the observation of some geometries not included in the main regions of MD distribution curves, as it may not represent distinct conformers but may represent average ones.

Although the correct reproduction of the oligosaccharide geometry is important, the flexibility associated with disaccharides exposed to different environments must also be taken into account. In this context, it is proposed that the number of possible conformations for the dihedral angles of a given disaccharide, when comprising a glycan moiety of a glycoprotein, is considerably reduced at its core compared to the distal ends.[5] This appears to be caused by the conformational restriction promoted by the surrounding carbohydrate residues, although its geometry is similar to the one present in the free carbohydrate.[2] In accordance with such a

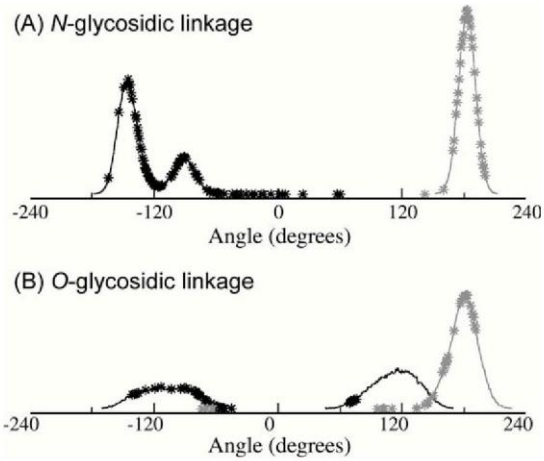


Figure 2. Distribution of the ϕ (black) and ψ (gray) dihedral angles associated with N- and O-glycosidic linkages in 50 ns MD simulations of the studied glycoproteins. Experimental geometries, as obtained from NMR data, are indicated by asterisks (*).

model, we were able to observe differences in the conformational behavior of the same disaccharide, β-GlcNAc-(1→4)-GlcNAc, in different molecular contexts (Figs. 1 and 4). In linear chains, with little or no steric hindrance, as seen in CD59 (PDB code 1CDR, Fig. 4A), a more flexible pattern is observed. In contrast, when such disaccharide is the core of a more complex glycan, as seen hsCD2$_{105}$ and α-hCG (PDB codes 1GYA and 1HD4, respectively, Fig. 4B), additional rigidity occurs, although the main peaks are highly similar, that is, a related set of the main conformational states. Such profiles seem to be influenced by the solvent accessibility around the disaccharides, as a higher solvent accessible surface (SAS) is observed in more flexible glycosidic linkages (Supplementary data). A similar influence of solvent exposure is observed for α-Man-(1→3)-Man depending on its localization at the glycan structure, that is, a given disaccharide will have higher flexibility when more exposed to solvent (Supplementary data). On the other hand, the presence of a linked carbohydrate residue alone does not necessarily influence the surrounding residues accessibility to the solvent. For example, the conformational pattern of the β-GlcNAc-(1→4)-GlcNAc linkage in the two CD59 PDB structures (1CDR and 1CDS) appears to not be influenced by a fucose six-linked to the GlcNAc residue (Supplementary data). This is probably due to the high flexibility of the (1→6)-linkage, which has less steric hindrance with the remaining glycan residues and, consequently, less influence on the solvation of the surrounding glycan.

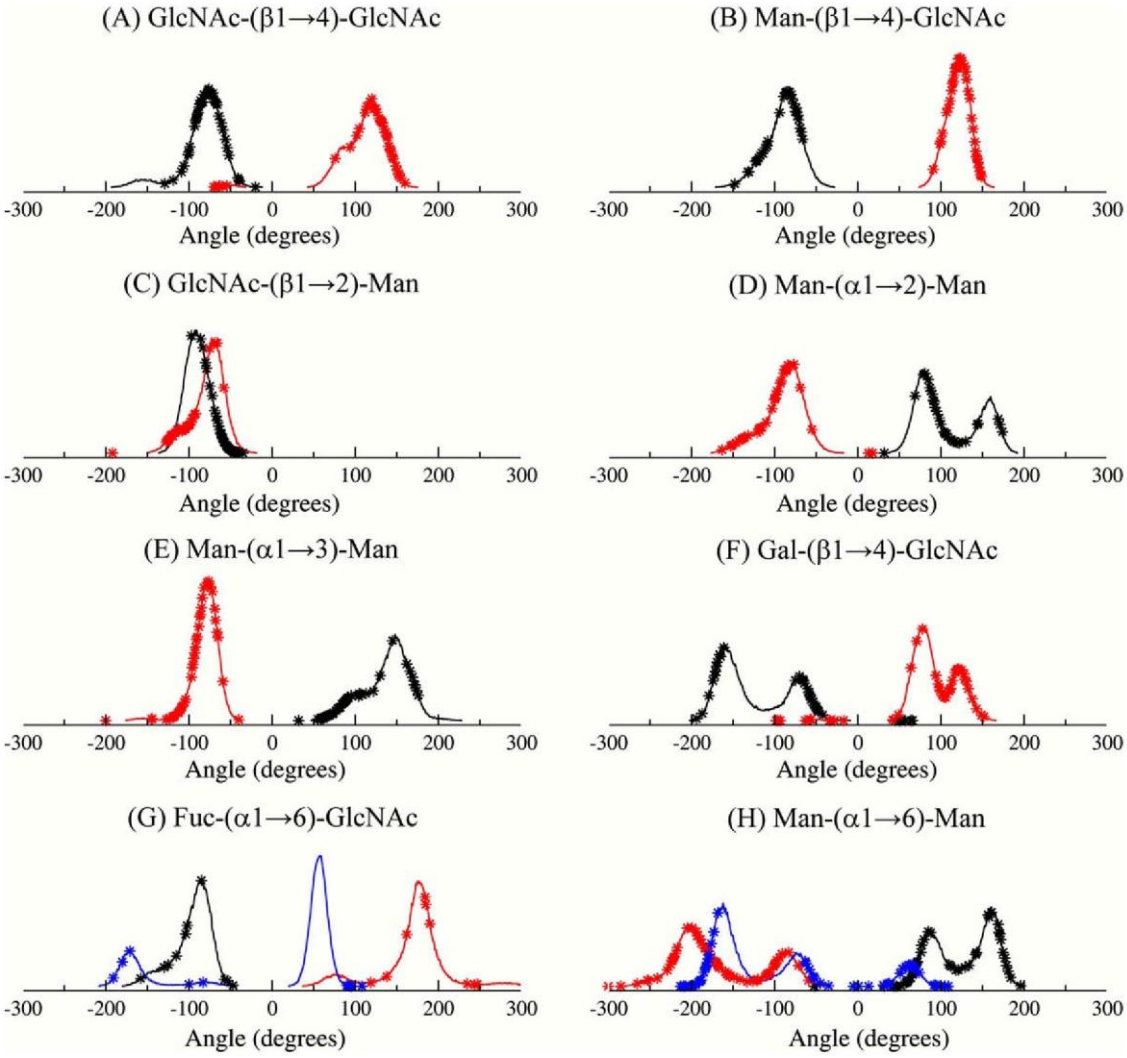


Figure 3. Distribution of the ϕ (black), ψ (red) and ω (blue) dihedral angles associated with specific disaccharides that compose the carbohydrate moieties of the studied glycoproteins in 50 ns MD simulations. Each curve is representative for the conformational behavior of a given disaccharide through the simulated glycoproteins (see Supplementary data). Asterisks (*) in the distribution curves indicate the experimental geometries, as obtained from NMR data, in a total of 836 values.

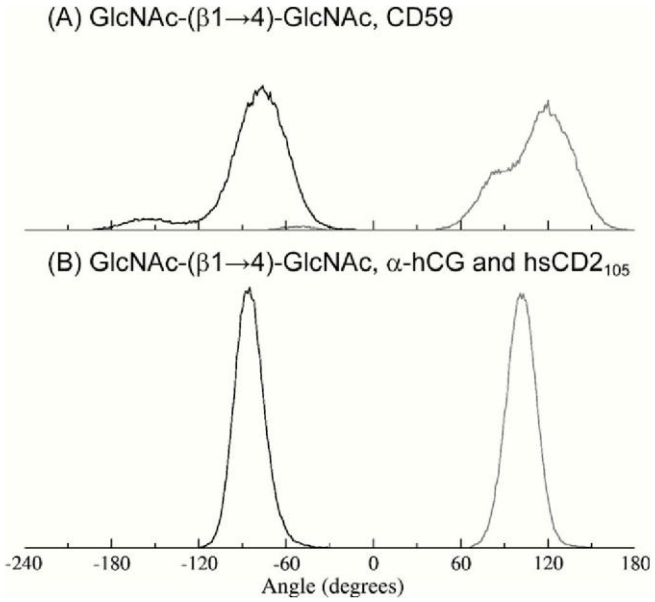


Figure 4. Distribution of the ϕ (black) and ψ (gray) dihedral angles associated with the β-GlcNAc-(1→4)-GlcNAc disaccharide behavior during 50 ns MD on different glycoproteins: (A) Attached to CD59 (PDB ID 1CDR and 1CDS); (B) Attached to both $hsCD2_{105}$ (PDB ID 1GYA) and α-hCG (PDB ID 1HD4).

Among the studies upon the conformational preferences of specific glycosidic linkages, special attention should be given to the description of the ω angle preferred orientations.[45] This bond is reported to adopt three main conformational states, around 180°, 60° and −60°, which is in good agreement with the conformational behavior observed from MD simulations. In contrast, it can be observed that NMR-measured dihedrals with regard to the ω angle of both α-Fuc-(1→6)-GlcNAc and α-Man-(1→6)-Man disaccharides adopt distinct geometries with large amplitudes (Fig. 3). These may indeed represent average values from two or three conformational states, which are observed to occur in solution for such disaccharides through MD.

3.4. Effect of glycosylation upon the protein moiety of glycoproteins

Although the biological functions of glycosylation are still not completely understood,[2,46] it is known that the glycan moieties of glycoproteins are fundamental to many biological processes, and are also involved in protein folding and/or stabilization.[1,2] In previous studies, MD simulations were employed to explore the role of glycosylation on peptides and proteins, such as the C-terminal region of the human prion glycoprotein,[47] the MUC1 peptide,[48] and a glycopeptide derived from a hemaglutinin protein fragment,[49] employing the AMBER force field,[7] and the MHC class I glycoprotein,[50] which used CVFF.[9] In this context, to evaluate simultaneously the glycan conformational description in MD, we have also performed an analysis of the influence of glycosylation on the dynamics and conformation of the protein moiety (Figs. 5 and 6), and these results will be discussed below. Additionally, the non-glycosylated molecules were also simulated with the OPLS-AA force field, with a secondary structure pattern equivalent to that observed with GROMOS96 (Supplementary data), which support the reliability of the obtained results.

3.4.1. EGF-like

According to experimental studies,[24] fucosylation does not significantly affect the structure of the EGF-like motif. In agreement with such an observation, no major differences in the secondary structure pattern between non-fucosylated and fucosylated proteins could be observed in MD simulations, as for RMSD (Fig. 5E) and radius of gyration (Fig. 5I). A high RMSD value for both forms of the EGF-like motif was observed, which appears to be due to the high loop content of this 46 amino acids protein, mainly its C- and N-terminal regions. On the other hand, a difference in flexibility between glycosylated and non-glycosylated proteins is observed around residues 56–59 (Fig. 6A and B), suggesting that the recognition of EGF-like domains by the *O*-fucosyltransferase may be associated with an entropic driven process, as this region lies within a putative consensus sequence for fucosylation in several EGF-like domains.[24]

3.4.2. CD59

The soluble form of CD59 had been studied by NMR methods, revealing a monomeric protein of 77 residues in a non-glycosylated form or with a disaccharide or trisaccharide N-glycan attached to Asn18.[21] Such data indicated that the global conformation of the protein backbone is essentially unchanged upon inclusion of oligosaccharides in the structure calculations. In accordance to these observations, MD simulations of non-glycosylated and glycosylated forms of CD59 showed no major differences between both forms of the protein for RMSD (Fig. 5F) and radius of gyration (Fig. 5J). Additionally, the NMR data pointed to small changes in the mean dihedrals and angular parameters of the protein component upon inclusion of carbohydrate groups, mainly around the side chain of Asn18.[21] Accordingly, the RMSF analyses of the simulated proteins support a decreased flexibility in the region of the N-glycan attachment, around Asn18 (Fig. 6C and D).

3.4.3. $hsCD2_{105}$

Regarding $hsCD2_{105}$, no major differences could be observed between non-glycosylated and glycosylated proteins based on RMSD and radius of gyration (Fig. 5G and K). However, an increased RMSD occurs in both forms of the protein, probably due to high loop content (mainly C- and N-terminal regions, Fig. 5C and G) and to the formation of several β-strands during MD simulations, known to exist in solution.[22,28,51] However, according to PROCHECK analysis, they are present in lower amounts in the NMR structures of both glycosylated and non-glycosylated forms (Fig. 5N).

Additionally, the NMR data suggest that the N-linked glycan at Asn65 is required for adhesion functions, because it stabilizes protein folding by counterbalancing an unfavorable clustering of five positive charges centered around Lys61.[22] Such an effect appears to be observed in MD simulations, because the region around these residues presents decreased flexibility upon glycosylation, as well as around regions comprising residues 22–32, 49–53, 61–67, and the N-terminal (Fig. 6E and F) domain.

Regarding CD2 function, transmembrane variants with mutations in the consensus N-glycosylation sequence Asn65-Gly66-Thr67,[52,53] which prohibit the attachment of the high-mannose N-glycan at Asn65, could be normally expressed on cell surfaces. However, they have shown neither antibody- nor ligand-binding activity, suggesting that the N-linked glycan on $hsCD2_{105}$ plays an important role in the function of this protein. In addition, the NMR studies suggested that protein–carbohydrate interactions may be involved in mediating the CD2 interaction with CD58 which should involve, mainly, the amino acids 32–34, 43–51 and 86–91 of $hsCD2_{105}$.[52,54] Accordingly, the residues around these regions concentrate the main differences in protein flexibility upon glycosylation (Fig. 6E and F), while participating in a series of β-sheets folded in MD apparently promoted by the glycan presence (Fig. 5N), suggesting a possible role of β-strand structures on CD2–CD58 binding.

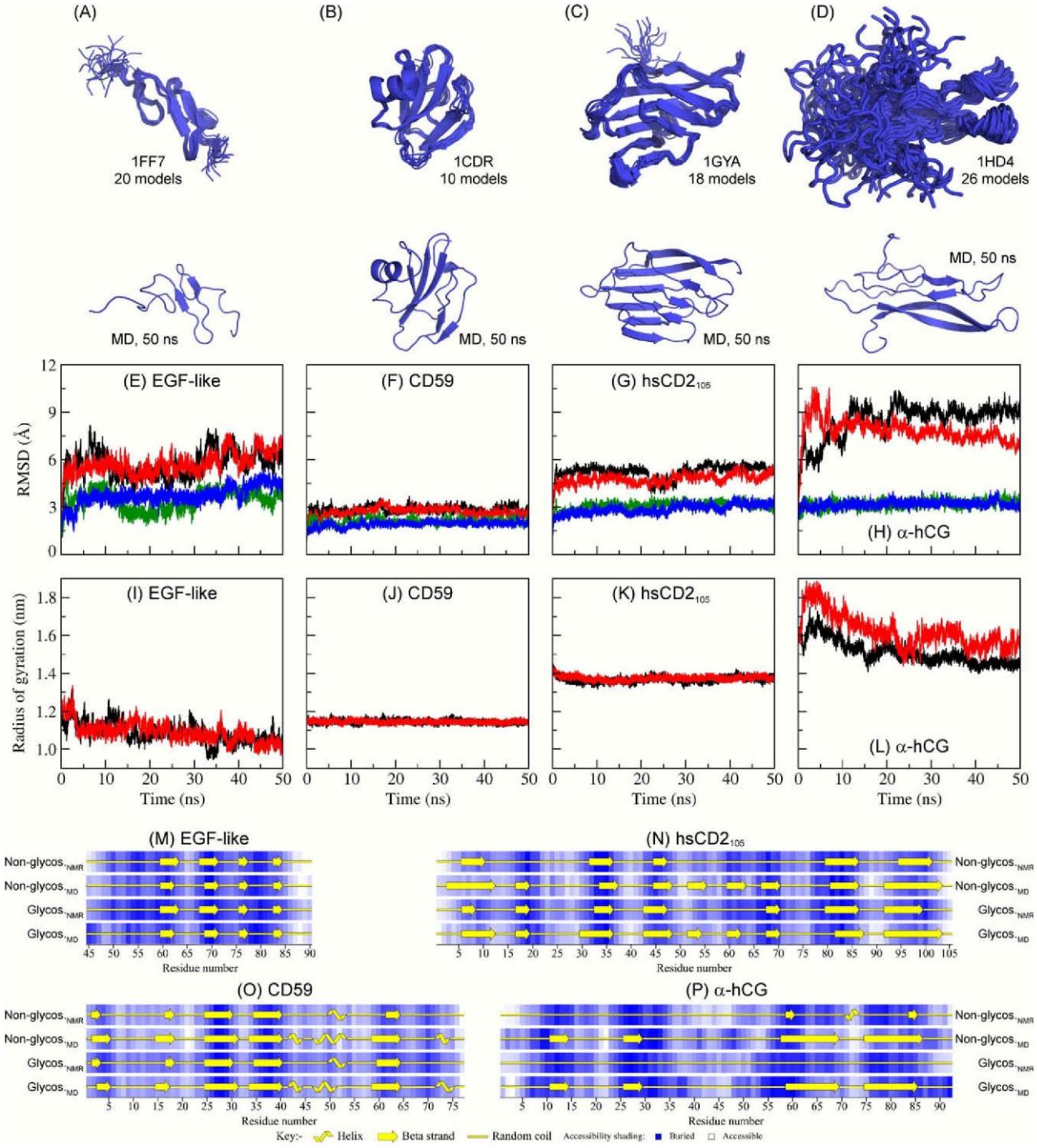


Figure 5. Comparative analysis between the studied glycoproteins and their non-glycosylated counterparts. The superimposed NMR models and the final MD conformation of EGF-like (A), CD59 (B), hsCD2$_{105}$ (C) and α-hCG (D) are presented together with root mean square deviation (RMSD) (E, F, G and H, respectively), in relation to the starting NMR model, for its respective non-glycosylated (black) and glycosylated (red) forms considering all atoms and the radius of gyration (I, J, K and L, respectively). The RMSD profile associated only with the atoms included in secondary structure elements of each molecule are shown for the non-glycosylated (green) and glycosylated (blue) forms. A comparative PROCHECK analysis of secondary structure content, obtained from NMR and final MD conformations of non-glycosylated and glycosylated proteins, is also shown (see Section 2 for further details).

3.4.4. α-hCG

The α-hCG structure has been determined through NMR methods in its glycosylated and non-glycosylated forms and deposited under PDB codes 1HD4[23] and 1DZ7,[27] respectively. Due to its high loop content, increased RMSD values are observed during MD simulations for both forms of the protein (Fig. 5H). This may also be related to the formation of a series of β-sheets in the simulations (Fig. 5P), which is also observed in OPLS/AA simulations of the non-glycosylated protein (Supplementary data). Additionally, some important differences could be observed between non-glycosylated and glycosylated forms of α-hCG based on RMSD (Fig. 5H) and radius of gyration (Fig. 5L). These differences seem to be caused by different loop flexibilities during the MD simulations (Fig. 5H), but especially by the large disordered loop between

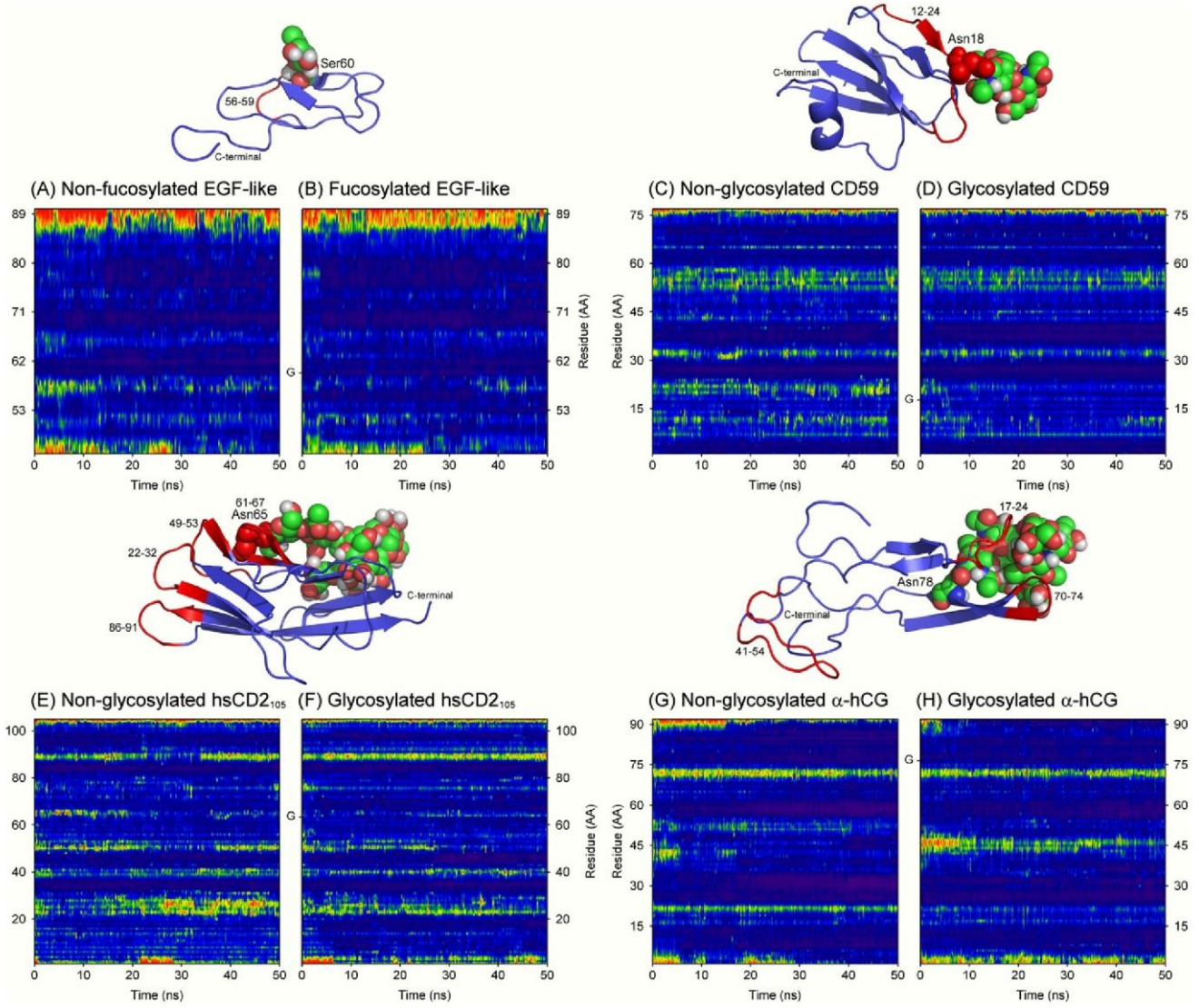


Figure 6. Root mean square fluctuation (RMSF) for the studied proteins, as a function of both residue number and time: EGF-like, in its non-fucosylated (A) and fucosylated (B) forms (from 0.04 to 0.24 nm); CD59, in its non-glycosylated (C) and glycosylated (D) forms (from 0.03 to 0.18 nm); $hsCD2_{105}$, in its non-glycosylated (E) and glycosylated (F) forms (from 0.04 to 0.16 nm); and the α-hCG, in its non-glycosylated (G) and glycosylated (H) forms (from 0.05 to 0.35 nm).

residues 33–58 (Fig. 5P). This loop is known to exist in free α-hCG, based on NMR studies,[55] and seems to adopt increased flexibility in the glycosylated protein (Fig. 6G and H). Such behavior is reproducible in an additional simulation, taking a different NMR model as starting structure under the same PDB code. As a general feature, glycosylation of the hCG α-subunit at Asn78 seems to play an important role in the stabilization of the protein[23] by exerting a protective function through shielding the protein surface from the environment,[55] including residues around 23–26 and 68–70. Therefore, from the MD simulations, we could observe a stabilization of such regions, especially 23–26, as observed in RMSF analysis (Fig. 6G and H).

3.5. Conformation of glycoproteins and their assessment through MD simulations

In recent years, glycobiology has become a critical facet of post-genomic science[56] as many proteins are post-translationally modified by glycosylation or co-translationally added by N-glycans in the endoplasmic reticulum.[46] Glycoslation may alter and regulate the biological activities of such macromolecules. In this context, several structural and functional studies on glycans alone, or on glycopeptides or glycoproteins have been carried out, using both experimental and theoretical methods. Concerning the use of the latter techniques, an assortment of approaches has been employed to model glycan structures when its three-dimensional structure is absent or is considered inappropriate. These include energy minimization of an entire oligosaccharide,[47] submitting the glycan to a simulated annealing protocol,[57] or constructing the carbohydrate moiety of glycoproteins from the disaccharide level,[50] that is, minimized structures of isolated disaccharides used to assemble an oligosaccharide tree. Additionally, several force field parameters have been employed to describe and simulate carbohydrate moieties of glycoproteins.[7–9] However, data regarding to the geometry of glycosidic linkages of such biomacromolecules are not always compared to experimental data[58] and, at times, are not analyzed during the simulated time,[47] that is, a proper populational ensemble characterization.

In contrast to most nucleic acids and proteins, which are linear and have a unique type of linkage, carbohydrates can be branched, linked through one of two anomeric configurations,[4] and there are more than 100 possible monomers (whereas ~10 in glycan composition),

which may be connected by different atoms, increasing its structural diversity.[59] All these properties, together with environmental factors, such as hydration and target proteins, determine the conformational ensemble of oligosaccharides. They are very flexible molecules, as their minimum-composing units, disaccharides, can adopt a dynamic equilibrium between two or more distinct conformations. All these properties are associated to the many functional roles of carbohydrates, making them one of the most challenging classes of molecules for conformational analysis.[6] Indeed, they are considered to have several orders of magnitude higher potential information content than any other biological macromolecule.[59] In this context, standard experimental techniques face several challenges when employed to obtain atomic-level structural information about carbohydrates. For example, X-ray crystallography may be difficult on these highly flexible systems and NMR spectroscopy mainly provides time-averaged conformational data.[6] Therefore, molecular modeling techniques, such as MD simulations, emerge as a promising tool for structural and conformational representation of carbohydrates, as they can describe, with a high level of accuracy in both spatial and temporal components, their geometry, flexibility and interaction with target proteins.[10–13,50,57,60,61]

Several force field parameters have been modified and/or designed to model and simulate carbohydrates, such as AMBER'93,[7] AMBER'06,[62] CVFF,[9] and CHARMm.[8] In this context, a previous paper pointed out that different approaches may be better for simulating different aspects of saccharides.[63] For example, the AMBER'93 force field seemed to be better than CHARMm for describing glycan dynamics and interactions with proteins, but less accurate for describing glycan hydration.[4] This may indicate that there is not a definitive force field for carbohydrates MD simulations, whereas further improvements were achieved with AMBER'06. On the other hand, this is the first systematic work assessing glycoprotein dynamics employing GROMOS96 43a1 force field parameters, together with the GROMACS simulation suite, which was shown to adequately represent the carbohydrate (and protein) moieties, as their NMR-derived conformations were adequately reproduced in a series of 50 ns MD simulations. As well, the simulation time scale achieved in this work totalized 0.4 μs for glycoproteins MD simulations and 0.2 μs for their non-glycosylated counterparts.

Although the modeling and adequate reproduction of glycans conformation are crucial for the adequate description of glycoproteins function, the analysis of N- and O-glycosidic linkages is essential. This may be because this relationship determines the orientation of the whole oligosaccharide relative to the protein and, in turn, influences the functional role of this class of biomacromolecules, such as the exposure of the glycan chain on the cell surface.[38] Among the MD studies on glycoproteins, only a few have explored the geometry of such linkages, including the α-GalNAc-(1→O)-Ser motif in solution,[64] and an antifreeze glycopeptide,[58] both consisting of O-glycosidic linkages with a carbohydrate moiety bound up to a disaccharide. The current work analyzed and properly described the conformations of an O-glycosidic linkage in a fucosylated glycoprotein, and the conformation of N-glycosidic linkages in four different glycoproteins. Among the parameters studied were its geometry and flexibility, indicating that the employed methodology is adequate for the reproduction of glycoprotein conformational ensembles in solution.

4. Conclusions

It is well known that glycoproteins play important roles in many biological events and that their carbohydrate moieties may be directly involved in such functions, including protein stabilization, folding, and interaction with target molecules. However, experimental methods present limitations when applied to carbohydrates. For example, X-ray crystallography is difficult on these highly flexible systems, and NMR tends to provide time-averaged conformations. Therefore, MD simulations emerge as a promising tool for complementing both NMR and X-ray experimental data on glycans and glycoproteins, because it provides a reasonable sampling of the conformational behavior of such molecules in solution.

The employment of molecular modeling techniques may be expected to contribute significantly to the understanding of the many events related to this class of biomacromolecules. Indeed, some force field parameters have been designed for modeling and simulating glycans, being shown to adequately represent experimental geometries. In this context, the GROMOS96 43a1 force field, enhanced by Löwdin HF/6-31G**-derived atomic charges for the glycan moiety, adequately reproduced the conformational ensemble of a series of glycoproteins, as well as the effect of glycosylation over the studied proteins, when compared to NMR data.

Finally, the employed methodology appeared to reproduce adequately the conformational ensemble of a series of glycoproteins containing carbohydrate moieties that are involved in different biological roles. Consequently, upcoming studies using MD simulations upon glycoproteins may be expected to provide a significant contribution to the investigation of their involvement in biological events.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq #420015/2005-1 and #472174/2007-0), MCT, and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2008.12.025.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
3. Bohne-Lang, A.; von der Lieth, C. W. *Pac. Symp. Biocomput.* **2002**, *7*, 285–296.
4. Petrescu, A. J.; Wormald, M. R.; Dwek, R. A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 600–607.
5. Petrescu, A. J.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **1999**, *9*, 343–352.
6. Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–216.
7. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832–3846.
8. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carbohydr. Res.* **1988**, *180*, 207–221.
9. Hwang, M. J.; Ni, X.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *Biopolymers* **1998**, *45*, 435–468.
10. Verli, H.; Guimarães, J. A. *Carbohydr. Res.* **2004**, *339*, 281–290.
11. Verli, H.; Guimarães, J. A. *J. Mol. Graphics Modell.* **2005**, *24*, 203–212.
12. Becker, C. F.; Guimarães, J. A.; Verli, H. *Carbohydr. Res.* **2005**, *340*, 1499–1507.
13. Pol-Fachin, L.; Verli, H. *Carbohydr. Res.* **2008**, *343*, 1435–1445.
14. 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 Zurich: Switzerland, 1996.
15. IUPAC-IUB Commission on Biochemical Nomenclature, *Pure Appl. Chem.* **1983**, *55*, 1269–1272.
16. Schaftenaar, G. MOLDEN. CAOS/CAMM Center, University of Nijmegen, Toernooiveld 1, 6525 ED NIJMEGEN, The Netherlands, 1997.
17. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
18. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
19. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

20. Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr., Sect. D* **2004**, *60*, 1355–1363.
21. Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D. *Structure* **1994**, *2*, 185–199.
22. Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A. R. N.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269*, 1273–1278.
23. Erbel, P. J. A.; Karimi-Nejad, Y.; van Kuik, J. A.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Biochemistry* **2000**, *39*, 6012–6021.
24. Kao, Y. H.; Lee, G. F.; Wang, Y.; Starovasnik, M. A.; Kelley, R. F.; Spellman, M. W.; Lerner, L. *Biochemistry* **1999**, *38*, 7097–7110.
25. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
26. Kräutler, V.; Müller, M.; Hünenberger, P. H. *Carbohydr. Res.* **2007**, *342*, 2097–2124.
27. Erbel, P. J. A.; Karimi-Nejad, Y.; De Beer, T.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1999**, *260*, 490–498.
28. Withka, J. M.; Wyss, D. F.; Wagner, G.; Arulanandam, A. R.; Reinherz, E. L.; Recny, M. A. *Structure* **1993**, *1*, 69–81.
29. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
30. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
31. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
32. Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
33. Jain, S.; Sundaralingam, M. *J. Biol. Chem.* **1989**, *264*, 12780–12784.
34. Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. *J. Mol. Biol.* **2005**, *351*, 431–442.
35. Andrec, M.; Snyder, D. A.; Zhou, Z.; Young, J.; Montelione, G. T.; Levy, R. M. *Proteins* **2007**, *69*, 449–465.
36. Imberty, A.; Pérez, S. *Protein Eng.* **1995**, *8*, 699–709.
37. Petrescu, A. J.; Milac, A. L.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R. *Glycobiology* **2004**, *14*, 103–114.
38. Ali, M. M.; Aich, U.; Varghese, B.; Pérez, S.; Imberty, A.; Loganathan, D. *J. Am. Chem. Soc.* **2008**, *130*, 8317–8325.
39. Hashimoto, Y.; Toma, K.; Nishikido, J.; Yamamoto, K.; Haneda, K.; Inazu, T.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1999**, *38*, 8377–8384.
40. Sticht, H.; Pickford, A. R.; Potts, J. R.; Campbell, I. D. *J. Mol. Biol.* **1998**, *276*, 177–187.
41. Pickford, A. R.; Smith, S. P.; Staunton, D.; Boyd, J.; Campbell, I. D. *EMBO J.* **2001**, *20*, 1519–1529.
42. Vakonakis, I.; Langenhan, T.; Prömel, S.; Russ, A.; Campbell, I. D. *Structure* **2008**, *16*, 944–953.
43. Hansen, A. P.; Petros, A. M.; Meadows, R. P.; Nettesheim, D. G.; Mazar, A. P.; Olejniczak, E. T.; Xu, R. X.; Pederson, T. M.; Henkin, J.; Fesik, S. W. *Biochemistry* **1994**, *33*, 4847–4864.
44. Barthe, P.; Pujade-Renaud, V.; Breton, F.; Gargani, D.; Thai, R.; Roumestand, C.; de Lamotte, F. *J. Mol. Biol.* **2007**, *367*, 89–101.
45. Shefter, E.; Trueblood, K. N. *Acta Crystallogr.* **1965**, *18*, 1067–1077.
46. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364–2369.
47. Zuegg, J.; Gready, J. E. *Glycobiology* **2000**, *10*, 959–974.
48. Rubinstein, A.; Kinarsky, L.; Sherman, S. *Int. J. Mol. Sci.* **2004**, *5*, 119–128.
49. Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421–8425.
50. Mandal, T. K.; Mukhopadhyay, C. *Biopolymers* **2001**, *59*, 11–23.
51. Wyss, D. F.; Withka, J. M.; Knoppers, M. H.; Sterne, K. A.; Recny, M. A.; Wagner, G. *Biochemistry* **1993**, *32*, 10995–11006.
52. Arulanandam, A. R. N.; Withka, J. M.; Wyss, D. F.; Wagner, G.; Kister, A.; Pallai, P.; Recny, M. A.; Reinherz, E. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11613–11617.
53. Recny, M. A.; Luther, M. A.; Knoppers, M. H.; Neidhardt, E. A.; Klandekar, S. S.; Concino, M. F.; Schimke, P. A.; Francis, M. A.; Moebiusll, U.; Reinhold, B. B.; Reinhold, V. N.; Reinherz, E. L. *J. Biol. Chem.* **1992**, *267*, 22428–22434.
54. Peterson, A.; Seed, B. *Nature* **1987**, *329*, 842–846.
55. de Beer, T.; van Zuylen, C. W. E. M.; Leeflang, B. R.; Härd, K.; Boelens, R.; Kaptein, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1996**, *241*, 229–242.
56. Turnbull, J. E.; Field, R. A. *Nat. Chem. Biol.* **2007**, *3*, 74–77.
57. Mukhopadhyay, C. *Biopolymers* **1998**, *45*, 177–190.
58. Nguyen, D. H.; Colvin, M. E.; Yeh, Y.; Feeney, R. E.; Fink, W. H. *Biophys. J.* **2002**, *82*, 2892–2905.
59. Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
60. Becker, C. F.; Guimarães, J. A.; Mourão, P. A. S.; Verli, H. *J. Mol. Graphics Modell.* **2007**, *26*, 391–399.
61. Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.* **1997**, *10*, 1249–1261.
62. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
63. Corzana, F.; Motawia, M. S.; Du Penhoat, C. H.; Pérez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. *J. Comput. Chem.* **2003**, *25*, 573–586.
64. Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. *J. Am. Chem. Soc.* **2006**, *128*, 14640–14648.

Supplementary Data

GROMOS96 43a1 performance on the characterization of glycoproteins conformational ensemble through molecular dynamics simulations

Laercio Pol-Fachin[a], Claudia Lemelle Fernandes[a] and Hugo Verli[a,b,*]

[a]*Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil*
[b]*Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av. Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil*

[*]Corresponding author. Tel.: +55-51-3308-7770; fax: +55-51-3308-7309; e-mail address: hverli@cbiot.ufrgs.br.

Table S1. NMR measured dihedral angle values for the glycan part of CD59 in the 10 models of PDB ID 1CDS.

PDB ID 1CDS	Dihedral angle (º)	
	GlcNAc(β1→4)GlcNAc	
	ϕ	ψ
Model 1	-119	-63
Model 2	-41	124
Model 3	-92	94
Model 4	-54	105
Model 5	-52	112
Model 6	-40	123
Model 7	-101	114
Model 8	-99	-70
Model 9	-70	-54
Model 10	-85	118

Table S2. NMR measured dihedral angle values for the glycan part of CD59 in the 10 models of PDB ID 1CDR.

PDB ID 1CDR	Dihedral angle (º)				
	GlcNAc(β1→4)GlcNAc		Fuc(α1→6)GlcNAc		
	ϕ	ψ	ϕ	ψ	ω
Model 1	-20	132	-125	162	-100
Model 2	-106	102	-104	-113	-171
Model 3	-74	-64	-102	-170	169
Model 4	-77	114	-57	-176	96
Model 5	-88	126	-47	-113	90
Model 6	-129	76	-158	119	108
Model 7	-90	-60	-130	-175	92
Model 8	-100	-69	-54	-116	94
Model 9	-56	120	-85	138	180
Model 10	-39	133	-115	-125	-83

Table S3. NMR measured dihedral angle values for the glycan part of $hsCD2_{105}$ in the 18 models of PDB ID 1GYA.

1GYA	Dihedral angle (°)																	
	GlcNAc(β1→4)GlcNAc		Man(β1→4)GlcNAc		Man(α1→3)Man		Man(α1→2)Man		Man(α1→6)Man			Man(α1→3)Man		Man(α1→6)Man			Man(α1→2)Man	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ω	ϕ	ψ	ϕ	ψ	ω	ϕ	ψ
Model 1	-85	137	-121	117	84	-87	100	-99	80	172	165	72	-105	158	-120	54	95	-92
Model 2	-82	141	-117	125	82	-82	95	-88	85	141	161	77	-65	171	-109	58	174	-69
Model 3	-92	126	-131	102	74	-78	106	-100	46	-169	180	101	160	112	-83	94	79	13
Model 4	-93	138	-114	113	84	-84	113	-140	52	172	-168	99	-117	121	-101	-45	169	-110
Model 5	-87	140	-117	116	79	-66	120	-118	68	-168	170	56	-94	143	-111	-62	95	-132
Model 6	-91	131	-123	107	74	-82	93	-121	60	-167	180	75	-92	157	-142	31	105	-143
Model 7	-86	133	-132	118	86	-90	107	-84	43	-149	178	110	-124	88	-133	38	130	-86
Model 8	-80	143	-111	123	84	-86	88	-88	71	136	152	156	-91	110	-176	109	81	18
Model 9	-82	136	-122	115	82	-76	92	-98	64	180	159	82	-69	142	-172	-5	122	-77
Model 10	-109	125	-119	99	88	-92	144	-96	76	167	-176	143	-119	72	-157	1	90	-97
Model 11	-89	142	-148	100	92	-96	108	-146	32	-156	-173	76	-144	-164	-156	153	91	-55
Model 12	-91	135	-119	121	77	-74	115	-115	43	-179	175	122	-110	161	-126	147	104	-151
Model 13	-83	131	-119	109	84	-85	107	-153	61	177	-179	134	-118	-176	-159	100	69	-89
Model 14	-91	129	-115	123	90	-93	90	-117	68	139	169	138	-97	134	-155	-58	87	-138
Model 15	-85	139	-108	111	79	-72	85	-85	47	176	162	125	-85	134	161	13	145	-93
Model 16	-99	131	-115	91	89	-88	98	-97	88	60	161	66	-145	129	166	-53	32	-163
Model 17	-80	135	-108	114	80	-68	116	-143	69	155	175	70	-124	77	-153	-35	171	-103
Model 18	-93	141	-121	111	86	-91	122	-136	64	-155	150	151	-104	118	171	61	113	-119

Table S4. NMR measured dihedral angle values for the glycan part of α-hCG in the 26 models of PDB ID 1HD4.

1HD4	Dihedral angle (°)																
	GlcNAc(β1→4)GlcNAc		Man(β1→4)GlcNAc		Man(α1→3)Man		GlcNAc(β1→2)Man		Gal(β1→4)GlcNAc		Man(α1→6)Man			GlcNAc(β1→2)Man		Gal(β1→4)GlcNAc	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ω	ϕ	ψ	ϕ	ψ
Model 1	-73	149	-82	129	61	-104	-35	-96	-171	106	66	178	-175	-35	-96	-171	106
Model 2	-73	146	-80	145	67	-76	-59	-120	166	64	179	-59	-59	-58	-106	-75	49
Model 3	-70	130	-74	139	174	-84	-62	-119	45	114	59	143	170	-63	-104	-47	122
Model 4	-70	115	-75	135	66	-108	-63	-126	-64	-28	59	143	-177	-67	-103	-165	42
Model 5	-69	151	-83	149	73	-110	-52	-122	-165	-93	67	95	49	-64	-104	-49	143
Model 6	-70	130	-72	143	63	-118	-48	-121	-59	138	161	137	-165	-66	-105	66	-34
Model 7	-76	149	-77	144	63	-112	-61	-112	51	125	92	-106	49	-97	-58	-48	152
Model 8	-71	131	-87	120	69	-40	-49	-126	49	129	177	-71	-44	-83	-69	49	119
Model 9	-66	130	-79	124	62	-113	-54	-125	-166	99	58	135	166	-75	-99	-49	122
Model 10	-83	152	-77	137	74	-97	-55	-127	-169	104	170	-121	170	-61	-106	-57	129
Model 11	-74	124	-94	121	66	-120	-49	-116	-70	-33	-179	-170	-51	-62	-102	-51	133
Model 12	-70	155	-89	130	55	-92	-54	-119	64	122	-52	107	170	-75	-92	-57	98
Model 13	-73	153	-69	146	171	-109	-42	-122	-166	-93	172	-126	-177	-59	-105	57	131
Model 14	-73	140	-109	119	62	-113	-40	-114	-56	121	175	-98	169	-79	-75	-161	-98
Model 15	-59	133	-69	147	172	-120	-43	-108	-60	-17	169	-100	65	-58	-104	52	120
Model 16	-66	139	-73	145	70	-127	-44	-124	-66	104	171	-91	49	-58	-95	-172	71
Model 17	-82	160	-73	133	68	-105	-55	-125	-55	110	51	75	-161	-79	-103	-65	-56
Model 18	-83	155	-68	131	61	-90	-51	-113	-166	102	64	177	174	-61	-103	-60	119
Model 19	-64	138	-66	141	68	-59	-47	-127	-65	-57	36	108	42	-57	-106	-55	149
Model 20	-93	152	-83	125	176	-118	-62	-127	-170	94	105	-97	68	-66	-102	-59	102
Model 21	-83	155	-74	142	58	-123	-53	-126	-174	138	-63	138	-140	-86	-66	173	112
Model 22	-90	160	-85	115	166	-95	-48	-104	-171	97	54	131	-59	-80	-93	-50	117
Model 23	-92	133	-77	130	60	-104	-64	-127	-54	-33	68	-111	-67	-60	-192	-52	128
Model 24	-67	145	-71	145	107	-86	-52	-100	-165	-96	172	-94	84	-61	-107	-169	111
Model 25	-61	132	-71	146	169	-90	-61	-97	-58	121	62	-108	-68	-72	-104	-179	78
Model 26	-71	126	-71	148	162	-118	-45	-121	-42	152	74	119	-64	-69	-102	-59	131

4.3 Trabalho II

Considerando a dificuldade de caracterização da estrutura e conformação de glicoproteínas, juntamente com suas cadeias sacarídicas correspondentes, e considerando a adequação dos procedimentos empregados pelo grupo na reprodução do equilíbrio conformacional de glicanas biológicas, o presente trabalho objetivou avaliar a adequação do emprego de estados conformacional de dissacarídeos, obtidos através de simulações de DM em solução aquosa, como guias para a construção de glicanas compondo glicoproteínas. Tal capacidade preditiva foi analisada a partir da comparação de dados conformacionais obtidos por DM para tais dissacarídeos com a geometria dessas ligações glicosídicas obtidas por RMN.

Os resultados obtidos indicam a co-existência de múltiplos confôrmeros para os dissacarídeos estudados em solução, especialmente para as ligações glicosídicas β1→4, de acordo com dados experimentais e teóricos prévios. Estes estados conformacionais mostraram-se semelhantes aos principais estados conformacionais populados pelas mesmas unidades dissacarídicas quando compondo estruturas glicoprotéicas. Tal observação sugere que o arcabouço protéico, ao menos nos casos estudados, parece não ser capaz de induzir o aparecimento de novos estados conformacionais na parte sacarídica. Por fim, modelos atomísticos para as estruturas glicosiladas completas das enzimas PGHS-1 e PGHS-2, até então ausentes, foram obtidos como aplicação da metodologia apresentada. Assim, os dados obtidos sugerem que a abordagem proposta pode se constituir em uma promissora estratégia na construção de estruturas completas de glicoproteínas, na ausência de dados experimentais prévios e, assim, contribuir no entendimento do papel destes glicoconjugados em sistemas biológicos.

GROMOS96 43a1 performance in predicting oligosaccharides conformational ensemble within glycoproteins

Cláudia Lemelle Fernandes, Liana Guimarães Sachett, Laercio Pol-Fachin, Hugo Verli

Carbohydrate Research, **2010**, *345*; 663-671

Carbohydrate Research 345 (2010) 663–671

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins

C. L. Fernandes [a], L. G. Sachett [a], L. Pol-Fachin [a], H. Verli [a,b,*]

[a] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil
[b] Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil

ARTICLE INFO

Article history:
Received 7 August 2009
Received in revised form 14 December 2009
Accepted 18 December 2009
Available online 28 December 2009

Keywords:
Glycoproteins
Molecular dynamics
Carbohydrates
Glycosidic linkage
GROMOS96

ABSTRACT

In previous work [Pol-Fachin, L.; Fernandes, C. L.; Verli, H.; *Carbohydr. Res.* **2009**, *344*, 491–500], we had demonstrated that GROMOS96 43a1 force field and Löwdin HF/6-31G** -derived atomic charges, adequately represent a glycoprotein's conformational ensemble in aqueous solutions, taking as the starting geometries NMR-determined structures. Based on such data, the present work intends to evaluate the use of the main solution conformations of isolated disaccharides, to build the carbohydrate moiety of glycoproteins, for which no previous experimental information is available. The observed results suggested that the entire glycoprotein scaffold appears unable to promote major modifications in the conformational behavior of glycosidic linkages. Additionally, when compared to energy contour plots, the results support the use of solution ensembles, to refine vacuum conformations of carbohydrate databases in the assembling of glycoproteins 3D structures. Finally, such approach is applied to build a full glycosylated model for COX-1 and COX-2 enzymes.

1. Introduction

Carbohydrates constitute a group of biological polymers of enormous structural and conformational complexity, presenting several possible monomeric units, adopting different conformational states, linked in diverse ways and presenting a variable degree of branching. However, while the identification of saccharide units being attached to proteins and their connectivity is experimentally accessible, their 3D information is not easily obtained, which increases the difficulties in elucidating carbohydrate functions in glycoproteins.[1–3]

Oligosaccharide 3D structures and conformation are mainly determined by their glycosidic linkage geometries, as a function of composition, linkage type, and solvation.[3] As a consequence, an approach usually employed to describe the conformational preferences of disaccharides consists in the construction of energy contour maps.[4–8] Based on the most stable geometries described by such plots, it is possible to infer the structure of a complete glycoprotein, in which the glycan moiety is built on a previously determined 3D protein structure.[2]

The potential contribution of such theoretic strategy may be estimated, considering the difficulties of some time-honored methods in describing the structure of glycans and glycoproteins. In X-ray crystallography, the oligosaccharide is usually incomplete, as its high flexibility may impair both the crystallization and acquisition of electron densities.[9–11] Even when these data are obtained, X-ray structures of carbohydrates are usually of poor quality, as demonstrated by a recent survey of PDB entries containing oligosaccharides, which suggests that about one-third of carbohydrate data comprise significant errors in relation to their stereochemistry, nomenclature, and consistency with electron density maps.[12,13] In NMR, one of the main challenges lies on the difficulties in resolving NOE signals, mainly due to the high flexibility of glycans, or to the systems size.[2,9,10] Such flexibility produces a complex conformational ensemble, with the possibility of co-existence of a number of simultaneous conformers in a solution.[14–17]

In this context, the current work intends to evaluate the adequacy of using main solution conformations of the isolated disaccharides, to build the carbohydrate moiety of glycoproteins, in the absence of previous experimental information on a carbohydrate's 3D structure and conformation. Accordingly, the obtained data are applied in building a complete model for COX-1 and COX-2 enzymes, in their fully glycosylated forms. Such strategy is expected to accurately represent glycoproteins in biological solutions, offering a refinement to the currently employed proceedings, and so contributing in the understanding of carbohydrates biological roles at the atomic level.

* Corresponding author. Tel.: +55 51 3308 7770; fax: +55 51 3308 7309.
E-mail address: hverli@cbiot.ufrgs.br (H. Verli).

0008-6215/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2009.12.018

2. Experimental

2.1. Computational methods

2.1.1. Nomenclature and topology

The analyzed disaccharide units represent carbohydrate moieties from NMR-determined glycoproteins: the human complement regulatory protein CD59 (PDB ID 1CDR and 1CDS),[18] the adhesion domain of human CD2 (PDB ID 1GYA)[19] and the α-subunit of the human chorionic gonadotropin (PDB ID 1HD4)[20] (Fig. 1). Based on the glycoproteins' data, eight different disaccharides, composed of *N*-acetyl-D-glucosamine (D-Glc*p*NAc), D-mannose (D-Man*p*), L-fucose (L-Fuc*p*), and D-galactose (D-Gal*p*), bounded by α/β(1→2), α(1→3), β(1→4), and α(1→6) linkages, were evaluated.

The IUPAC nomenclature recommendations and symbols[21] have been adopted. The relative orientation of a pair of contiguous carbohydrate residues was described, for different types of linkages, by two or three torsional angles at the glycosidic linkage. For a (1→*X*) linkage, where '*X*' is '2', '3', '4', or '6' for the (1→2), (1→3), (1→4), or (1→6) linkages, respectively, the ϕ and ψ were defined as shown below:

$$\phi = \mathrm{O5{-}C1{-}OX{-}CX'}$$

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

For a (1→6) linkage, the ω was defined as

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

Each of the disaccharide structures was built in Molden[22] and their respective topologies were obtained from the PRODRG server.[23] The HF/6-31** -derived Löwdin charges were employed.[24,25] Improper dihedrals were included to preserve the hexopyranose conformations in accordance with their expected form in aqueous solution: $^4C_1$ for D-Glc*p*NAc, $^1C_4$ for L-Fuc*p*, $^4C_1$ for D-Man*p*, and $^4C_1$ for D-Gal*p*. These topologies were employed in molecular dynamics (MD) simulations and were analyzed using the GROMACS simulation suite[26] and GROMOS96 43a1 force field,[27] as described previously.[17,24,25,28,29]

2.1.2. Calculation of energy contours maps

The conformational description of each disaccharide was performed through a calculation of relaxed contour maps around the glycosidic linkages, rotating the torsion angles from −180° to 180° in steps of 30°, with a total of 144 conformers for each linkage. This was performed using a 334.8 kJ mol^{-1} constant force, to make a restriction only to the ϕ and ψ proper dihedrals during energy minimization, in each of the above-mentioned values, thus allowing the search of the conformational space associated with the disaccharide. Then, using the minimized output conformation,

(A) Human complement regulatory protein CD59 (PDB ID 1CDS)

β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-*Asn18*

(B) Human complement regulatory protein CD59 (PDB ID 1CDR)

β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-*Asn18*
6
↑
1
α-L-Fuc*p*

(C) Adhesion domain of human CD2 (PDB ID 1GYA)

α-D-Man*p*
1
↓
3
α-D-Man*p*-(1→2)-α-D-Man*p*-(1→6)-α-D-Man*p*
1
↓
6
β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-*Asn65*
3
↑
1
α-D-Manp-(1→2)-α-D-Manp

(D) Human chorionic gonadotropin (PDB ID 1HD4)

β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→2)-α-D-Man*p*
1
↓
3
β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-*Asn78*
6
↑
1
β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc-(1→2)-α-D-Man*p*

Figure 1. Schematics of the NMR glycan structures used to construct the simulated disaccharides.

a series of MD simulations were performed for 20 ps at 10 K, with an integration step of 0.5 fs to further support the search for minimum-energy conformations. In this process, the rotatable exocyclic groups were allowed to freely search for minimum-energy orientations.[29,30] The relative stabilities of each conformation, obtained from the 10 K MD simulations, based on GROMOS96 force field[27] and HF/6-31G** Löwdin atomic charges, were used to construct relaxed energy contour plots, describing the conformation of each glycosidic linkage, as shown in Figure 2.

2.1.3. MD simulations

The minimum-energy conformations of each of the disaccharides, as obtained from their contour plots, were then used as the starting conformations for MD simulations in an aqueous solution, using the GROMACS simulation suite.[26] Each pair of angles describing a vacuum minimum-energy conformation was used to build the disaccharide conformers that would be further submitted to MD simulations. The disaccharides were solvated in a triclinic box using periodic boundary conditions and SPC water model.[31] The Lincs method[32] 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 using the Particle Mesh Ewald method.[33] Temperature and pressure were kept constant by coupling disaccharides, ions, and solvent to external temperature and pressure baths, with coupling constants of $\tau = 0.1$ and 0.5 ps,[34] respectively. The dielectric constant was treated as $\varepsilon = 1$, and the reference temperature was adjusted to 310 K. The systems were slowly heated from 50 to 310 K, in steps of 5 ps, each one increasing the reference temperature by 50 K. Each simulation was extended to 0.1 μs.

2.1.4. Glycosylated cyclooxygenase (COX) models

The murine COX-2 and ovine COX-1 were retrieved from PDB IDs 1CVU and 1Q4G, respectively. These structures were isolated to their monomeric forms, in which, the N-glycosylation sites (Asn68, Asn144, and Asn410) were filled with a putative model for their glycan moieties, represented in Figure 4, using glycoscience modeling tools.[35–37] This model was proposed based on a consensus between previous biochemical, crystallographic and mass spectrometry data.[38,39] The composition evidenced by spectrometric results was compared to structures deposited in PDB and further refined by biochemical evidence, supporting the identification of a proper anomeric state for each residue and type of linkage for carbohydrate residues. The PDB structures were also submitted to pdb-care software,[40] in order to identify and correct

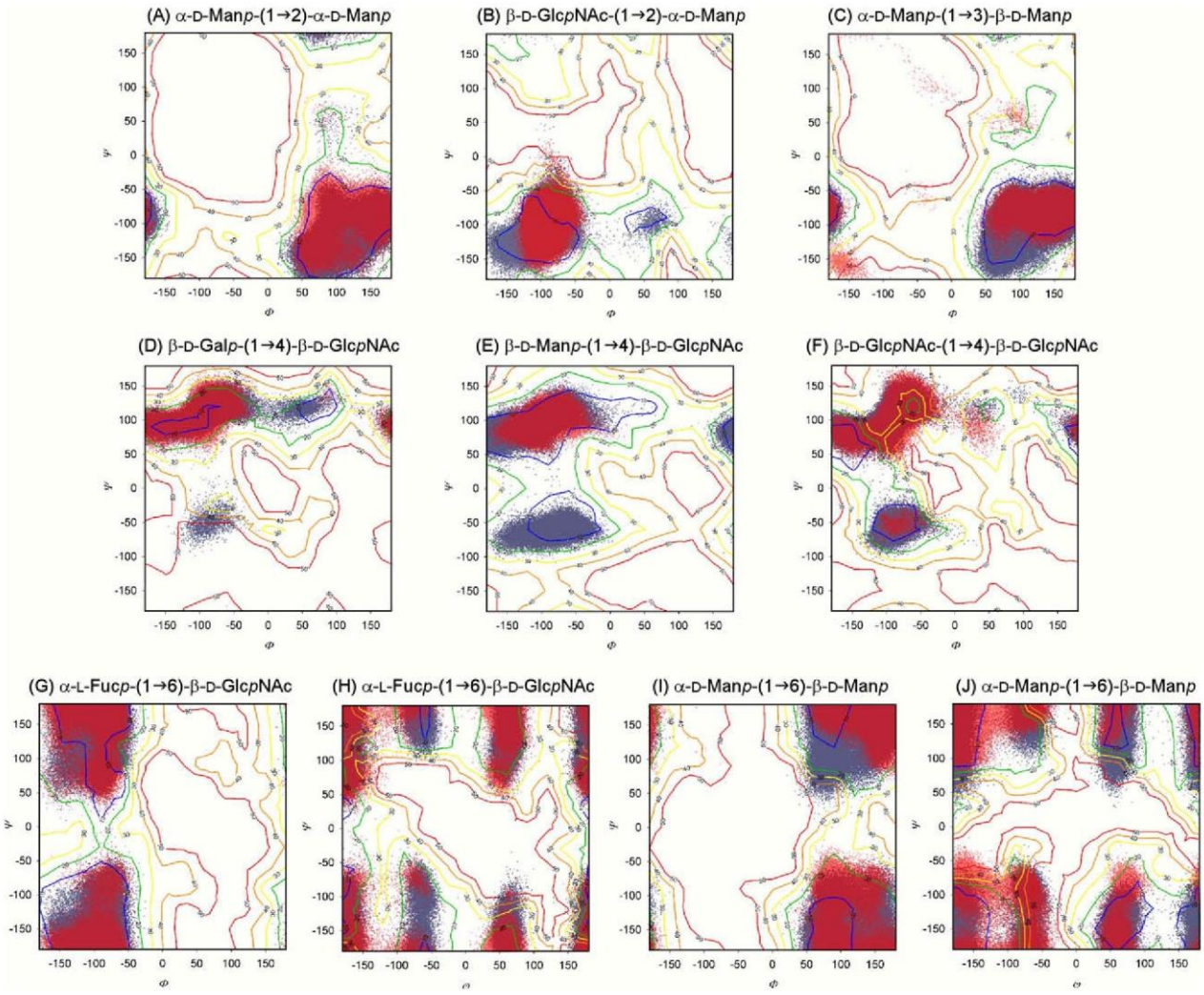


Figure 2. Relaxed contour plots for the studied glycosidic linkages, in which the energy maps are shown, every 10 kJ mol^{-1}, from 10 to 50 kJ mol^{-1}, superimposed with the results of MD simulations in solution. Blue dots indicate the conformational behavior of disaccharide units, while red dots indicate the conformational behavior of a given disaccharide in a whole glycoprotein environment.

potential structural errors. These so-obtained models for glycosylated COX-1 and COX-2 had their glycosidic linkage geometries adjusted to the main conformations for each linkage, based on their relative abundance in the isolated disaccharides in water, as described in Table 1. Such models were then submitted to 50 ns MD simulations in an aqueous solution, using the GROMACS package[26] and GROMOS96 43a1 force field,[27] following the simulation proceedings described above (Section 2.1.3). As previous results identified that detergent molecules do not appear to be essential to stabilize the membrane-binding domain during simulations[41] and that there is no glycosylation site within this domain, then the proteins were simulated without restraints in order to observe the dynamics of the catalytic and EGF domains, which in turn are glycosylated.

3. Results and discussion

3.1. Simulation systems

The strategy employed to evaluate the adequacy of using conformations of disaccharides, derived from MD as guiding geometries, for building the glycan parts of glycoproteins without experimental restraints from NMR, or X-ray crystallography, was based on the following steps: (1) building of disaccharides commonly found in glycoproteins; (2) calculation of their vacuum contour plots in order to obtain minimum-energy conformations, which were to be further sampled in solution; (3) MD simulations of each disaccharide's minimum-energy conformation in aqueous solutions, under an explicit solvent model; and (4) to compare the obtained conformational samplings, in both vacuum and solution, to previous MD and NMR data of glycoproteins. The studied disaccharides included: α-D-Man*p*-(1→2)-α-D-Man*p*, α-D-Man*p*-(1→3)-β-D-Man*p*, α-D-Man*p*-(1→6)-β-D-Man*p*, β-D-Glc*p*NAc-(1→2)-α-D-Man*p*, β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc, β-D-Man*p*-(1→4)-β-D-Glc*p*NAc, β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc, and α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc. The so-obtained data were then applied to describe the complete glycosylated structures of COX isoforms, for which until now, there were no 3D data to illustrate the application of the proposed protocol.

3.2. Vacuum conformational analysis

There is a considerable amount of data on disaccharide conformational preference in carbohydrate databases.[35–37,42] The Glycosciences portal[35–37] has a large amount of data based on MM3[43,44] force field, thus offering a highly detailed representation of carbohydrates conformational features,[45,46] at the least, in gas phase.[47,48] Unfortunately, such a force field has not been as widely used, to study biological systems, as more specifically parameterized force fields, such as CHARMM, AMBER, GROMOS, and OPLS, amongst others.[47,49] In this context, we proceeded to the calculation of relaxed maps for a series of eight disaccharides (Fig. 2), fre-

Table 1
Conformational preferences of disaccharides as observed from MD simulations[a]

Disaccharide	Relative abundances (%)	Angle (°)		
		ϕ	ψ	ω
α-D-Man*p*-(1→2)-Man*p*	**62**	**80 ± 25**	**−135 ± 36**	—
	38	160 ± 20	−100 ± 28	—
β-D-Glc*p*NAc-(1→2)-Man*p*	**96**	**−90 ± 35**	**−95 ± 32**	—
	4	60 ± 25	−100 ± 24	—
α-D-Man*p*-(1→3)-Man*p*	33	80 ± 24	−85 ± 28	—
	23	−130 ± 18	—	
	44	**160 ± 21**	**−80 ± 21**	—
β-D-Gal*p*-(1→4)-Glc*p*NAc	1.5	**−80 ± 29**	−50 ± 23	—
	86	**120 ± 21**	—	
	3	50 ± 29	—	
	10	−150 ± 17	90 ± 16	—
β-D-Man*p*-(1→4)-Glc*p*NAc	23	−70 ± 30	−55 ± 19	—
	33	110 ± 22	—	
	44	**−160 ± 21**	**90 ± 18**	—
β-D-Glc*p*NAc-(1→4)-Glc*p*NAc	**56**	**−80 ± 24**	**115 ± 16**	—
	19	−90 ± 28	−60 ± 20	—
	15	−100 ± 18	90 ± 14	—
	11	−150 ± 17	—	
α-D-Fuc*p*-(1→6)-Glc*p*NAc	14	**−90 ± 35**	120 ± 27	60 ± 16
	16	−160 ± 32		
	12	160 ± 34	−60 ± 17	
	3	−90 ± 24		
	17	90 ± 23	**−180 ± 17**	
	38	**180 ± 46**		
α-D-Man*p*-(1→6)-Man*p*	**27**	**80 ± 23**	**−160 ± 30**	**60 ± 17**
	6		110 ± 29	
	6	150 ± 18	−160 ± 31	
	7		110 ± 24	
	5	80 ± 22	−90 ± 21	−70 ± 19
	6		170 ± 34	
	1.5	150 ± 19	−90 ± 23	
	5		170 ± 30	
	11	80 ± 24	−125 ± 21	175 ± 19
	11		−180 ± 24	
	4		90 ± 21	
	3	150 ± 19	−125 ± 23	
	6		−180 ± 24	
	1		90 ± 20	

[a] The most abundant conformations are marked as bold and were used, as necessary, to construct the fully glycosylated COX enzymes models (see details in the text).

quently found in glycoproteins, under GROMOS96 43a1 force field set of parameters. The eight disaccharides analyzed provided 10 maps, including ϕ–ψ and ω–ψ plots.

Similarities between the conformational patterns of α-D-Man*p*-(1→2)-α-D-Man*p* and α-D-Man*p*-(1→3)-β-D-Man*p* may be observed, as reported,[50–52] with one broad minimum energy region in both ϕ and ψ dimensions (Fig. 2A and C). Likewise, the β(1→4) linkages are presented from two to three minimum energy regions, being in the northwest region of the plot, from β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc, β-D-Man*p*-(1→4)-β-D-Glc*p*NAc, and β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc, and also in the southwest quadrant for the latter two disaccharides. While the inversion in stoichiometry at carbon atoms 2 and 4 in Gal, regarding to Man, appears to increase the rigidity of the linkage, the β(1→4) disaccharides are globally more flexible than the α(1→2) and α(1→3) linkages, presenting more than one single energy minima, in agreement with the previous data.[53,54] Additionally, the conformational restriction promoted by the presence of *N*-acetyl group in position C-2, maintaining the location of the minima, was already described.[55]

Concerning the behavior of (1→6) linkages, the preference of the ω dihedrals in the region of −60°, 60°, and 180°, usually reported in the literature,[56–59] was observed accordingly for both α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc and α-D-Man*p*-(1→6)-β-D-Man*p* disaccharides (Fig. 2H and J, respectively). Furthermore, at least in vacuum, the steric hindrance promoted by the methyl group, at Fuc carbon 5 and the *N*-acetyl group at GlcNAc carbon 2, reduced the extension and the depth of minimum energy regions when compared to Man-containing disaccharides. As a general feature, all disaccharides presented well-defined minimum energy regions, in agreement with those deposited in the Glycosciences portal,[35–37] indicating a similar conformational description of the main energy minima between MM3 and GROMOS96 43a1 force fields,[27] for the evaluated systems. Therefore, such results indicate that the latter may be used in carbohydrates conformational description with a reasonable level of accuracy.

3.3. Solution conformational sampling

Considering that the explicit inclusion of solvent molecules is already described as able to reveal a distinct set of conformers when compared to calculations in its absence,[60] each energy minima from maps (Fig. 2) was submitted to 100 ns MD simulations in order to obtain a description of the solvent influence over each disaccharide (Supplementary data). In fact, such influence may be initially observed on the use of different energy minima, as starting geometries, for distinct sets of simulations. Independent of the starting minima, the obtained conformational ensembles were equivalent, suggesting a major role of solvent on a disaccharide's conformational equilibrium. Also, it indicates the capability of the employed time scale, to effectively describe the conformational transitions, relevant to each system.

As presented in Figure 2 and Table 1, the conformational preference of most of the simulated disaccharides remained located around their vacuum minimum energy regions. However, upon inclusion of a solvent, some less favorable vacuum regions were also occupied, adjacent to minimum energy regions, as well as some high energy regions from the vacuum map, indicating the influence of the solvent in the stabilization of additional regions of the energy maps. For example, the two energy minima shown by the β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc disaccharide in vacuum, increased to three main regions in the conformational space populated in solution (Fig. 2D and Table 1). Thus, it became more similar to the remaining β(1→4) simulated disaccharides (Fig. 2E and F), as expected.[36,54,55] Similarly, the α-D-Man*p*-(1→6)-β-D-Man*p* presents an additional stabilization of the ω angle at −60° as a consequence of solvent inclusion (Fig. 2J), acquiring a conformational preference similar to that shown by α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc (Fig. 2H).

On the other hand, some minimum energy regions on the contour plots were not consistently populated in solution, as could be observed for β-D-Glc*p*NAc-(1→2)-α-D-Man*p* (Fig. 2B) and β-D-Man*p*-(1→4)-β-D-Glc*p*NAc (Fig. 2E). These observations indicate that, while the vacuum conformational maps indeed generated good starting points for the representation of carbohydrates in solution, the conformational space described by the two strategies is not necessarily equivalent. Most importantly, such data reinforce the key importance in using explicit solvent effects when computationally studying biological carbohydrates and glycoconjugates.

3.4. Accuracy in disaccharide use in solution as a conformational reference for building models of glycans in glycoproteins

In spite of the fact that energy contour plots, in the absence of an explicit solvent, are well described in the literature under different force fields and usually having their minima set according to crystallographic data,[35] the direct association of these geometries, to infer solution conformations and ensembles of biological carbohydrates, has some important drawbacks: (1) the high amount of errors in the PDB of carbohydrate containing structures;[12,13] (2) usually, only a few carbohydrate residues are found in glycoproteins crystallographic structures, close to the protein core,[9,10] impairing the acquiring of information about carbohydrates conformations in the extremity of glycans; (3) aqueous solution absent environments tend to reveal a distinct set of conformers when compared to solution data;[60] (4) crystal packing effects promote both glycosidic linkage conformational modifications and hexopyranose ring distortions.[61] In this context, additional strategies capable of supplying carbohydrate and glycoconjugate conformational ensembles of biological significance may strongly contribute to the understanding of the functional roles of such molecules in living organisms.

Employing the GROMACS[26] package as a fast, low cost, simulation package, supporting the achievement of high conformational samplings, we had previously demonstrated that the GROMOS96 43a1 force field,[27] supplemented by Löwdin HF/6-31G** -derived atomic charges and PRODRG[23] topologies, for carbohydrate residues, is capable of adequately representing glycoprotein conformational ensemble in aqueous solutions, using the NMR data[17] as a reference ensemble. As this work employed the NMR data from PDB as the starting geometries, its results were compared to those obtained in the current work, where no previous experimental data were used to set the carbohydrate simulations, in order to check the reliability of disaccharide solution ensembles to reproduce the conformational behavior of equivalent units in complex glycans as glycoproteins.

Therefore, the conformational distribution of a set of glycosidic linkages was compared between its isolated disaccharides and also when within protein-bounded glycans, as shown in Figure 3. Besides, each glycosidic linkage conformation, as described in the employed NMR models from PDB, was included in the distribution curves as asterisks, in order to offer an experimental reference of each disaccharide conformational ensemble, in a total of 836 geometries. From these points, at least 90% were populated in each disaccharide, indicating the high reliability in simulating disaccharides to represent biological ensembles of glycoproteins as determined by NMR methods.

From all 18 dihedral angles analyzed, five presented virtually identical behaviors between the disaccharide and the glycan forms of the linkages, as for the ϕ angle from β-D-Glc*p*NAc-(1→2)-α-D-Man*p*, the ϕ angle from α-D-Man*p*-(1→2)-α-D-Man*p*, the ϕ and ψ angles from β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc, and the ϕ angle

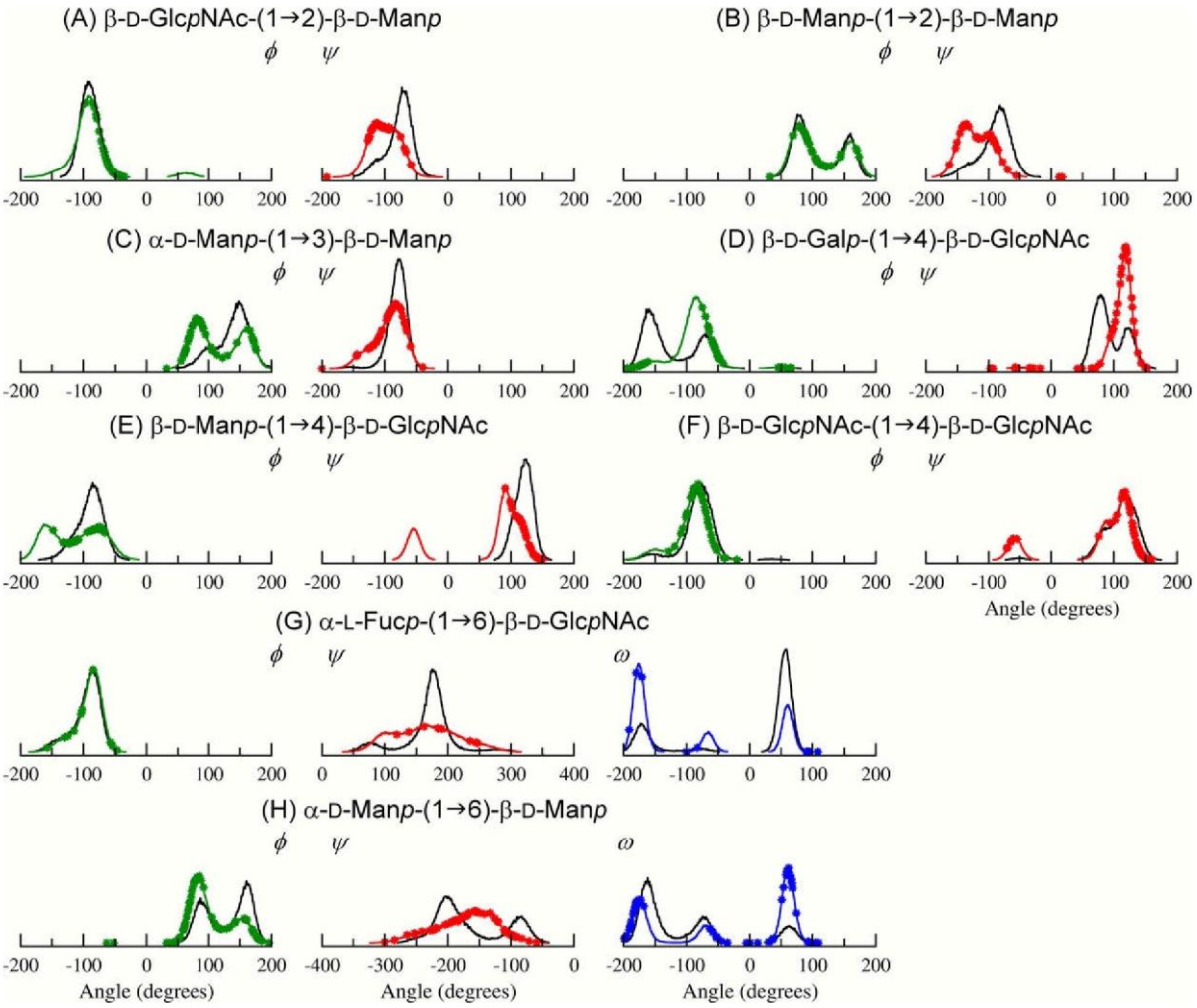


Figure 3. Distribution of the ϕ (green), ψ (red), and ω (blue) dihedral angles associated with the studied disaccharides in 0.1 μs MD simulation. The behavior of such glycosidic linkages in their protein-bounded glycans is also supplied (black curves). The asterisks ( ) in the isolated disaccharides distribution curves indicate the NMR experimental geometries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc, which suggests an absence of influence of the entire glycoprotein scaffold on specific disaccharide units, at least on the cases studied here. On the other hand, the six angles presented, as expected, had an increased flexibility in the isolated disaccharides when compared to the same units in the glycoprotein structure, suggesting some degree of conformational susceptibility to the chemical neighborhood, including the ψ angle from β-D-Glc*p*NAc-(1→2)-α-D-Man*p*, the ψ angle from α-D-Man*p*-(1→2)-α-D-Man*p*, the ψ angle from α-D-Man*p*-(1→3)-β-D-Man*p*, the ϕ and ψ angles from β-D-Man*p*-(1→4)-β-D-Glc*p*NAc, and the ψ angle from α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc. Curiously, both angles from β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc presented an inverse behavior, with the isolated disaccharides presenting a lower flexibility than any of the glycans, suggesting that the glycoprotein scaffold may be capable of shifting the conformer populations equilibrium for some glycosidic linkages. Nevertheless, the simulation of a glycoprotein, with its β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc, in the main conformation obtained from the isolated disaccharides, does offer the same conformational pattern as described by the NMR.[17] This indicates the plasticity of these linkages and the capability of MD simulations in explicit solvent, to adequately represent the expected conformational transitions of a given disaccharide unit in two different environments. Furthermore, in most of the cases where the glycosidic linkage geometries were not identical between the isolated disaccharide and the entire glycoproteins, there were no formations of new conformational peaks, but a modification in the relative abundance of the same conformational state. The major exception was the ψ angle from α-D-Man*p*-(1→6)-β-D-Man*p*, in which the main conformational state in the isolated disaccharide resembled an average conformation between the two conformational states observed in the same disaccharide when inserted in a glycoprotein molecular context.

In order to illustrate the potential of using disaccharide solution conformations as the starting geometries for building complete glycoproteins, we employed the main conformational states described in Figure 3 and Table 1, to build models for the complete glycosylated COX-1 and COX-2 enzymes. In fact, there are about 30 crystallographic structures deposited in PDB for the two proteins, most of it glycosylated (Supplementary data). Unfortunately, the majority of such structures present a small number of carbohydrate residues, that is, half of all crystal structures has up to two residues in each glycosylation position. Also, the structures presenting higher glycosylation patterns present only three Man residues, and for a single position, mostly Asn144 (PDB IDs 1IGX, 1IGZ, 1FE2, 1DIY, 1Q4G, and 2AYL), while 5–10 Man are expected to occur in all positions.[39,40] Most importantly, several errors have been observed in these structures (Supplementary data), as evidenced by pdb-care[41] and based on the expected glycan structure assem-

bled in the endoplasmic reticulum.[2,62] In this context, and considering the great therapeutic impact of cyclooxygenase inhibitors, a putative model was built for the glycan moiety of such enzymes (Fig. 4, see Section 2 for details), in agreement with most of the available experimental data on this system.[39,40] Such a glycan was employed to glycosylate both COX-1 and COX-2 in positions Asn68, Asn144, and Asn410.[63] The obtained glycoproteins were further submitted to MD simulations in order to refine the models and allow for carbohydrate accommodation in the presence of solvent and protein scaffold. A dimeric glycosylated COX-2 is presented in Figure 5, as well as a representation of the glycan ensemble through the simulations.

Each dihedral from glycosidic linkages was analyzed and compared to the conformational pattern observed for the proper disaccharide units (Fig. 4 for COX-2 and in Supplementary data for COX-1). Accordingly, as a general feature and previously pointed in the

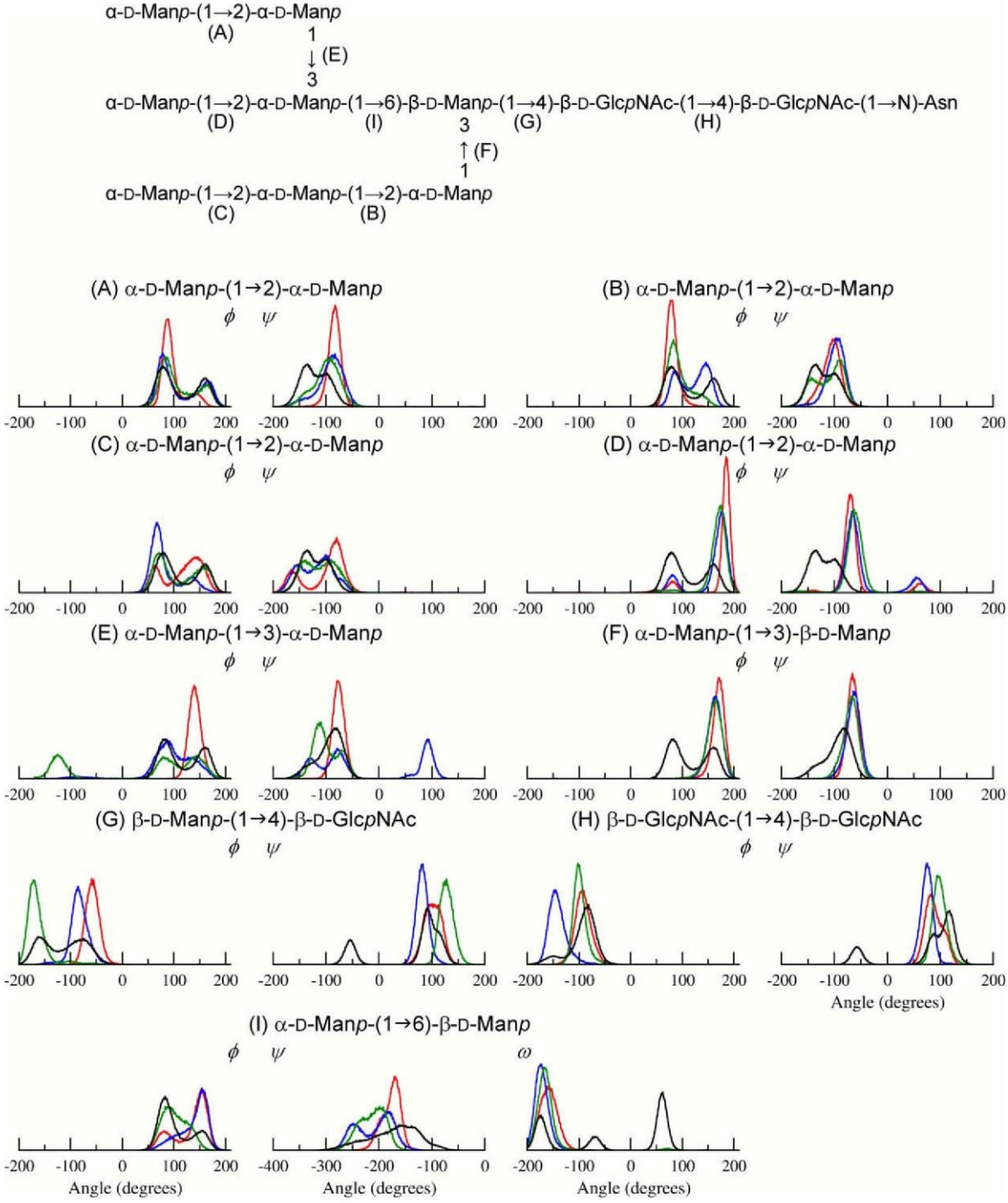


Figure 4. Distribution of the ϕ, ψ, and ω dihedral angles, as obtained from MD simulation, from COX-2 glycans at Asn68 (red), Asn144 (blue), and Asn410 (green), together with the conformational patterns associated with the respective disaccharides, isolated in water (black). The A–I letters indicate the relation of a distribution plot to its location on the respective disaccharide unit from the complete glycan. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

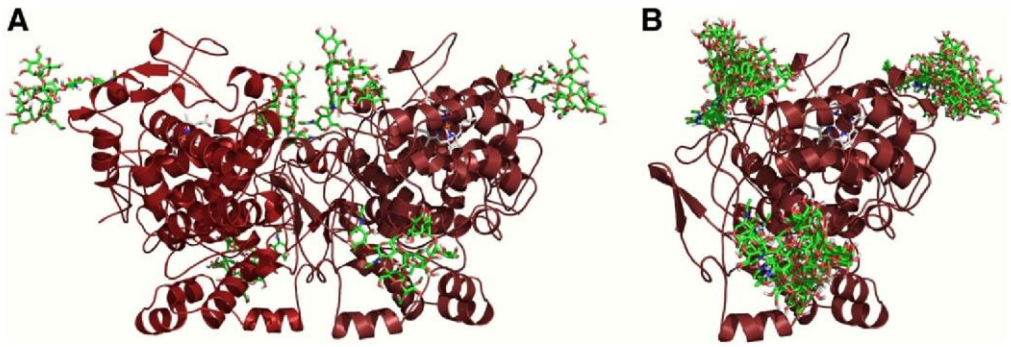


Figure 5. Glycosylated structure from murine COX-2 built based on the main conformational states as presented in Figure 3 and Table 1. (A) glycosylated COX-2 dimer and (B) superimposition of snapshots obtained from 50 ns MD of the monomeric fully glycosylated COX-2.

text, most of the linkages do not present new conformer populations, but distinct preferences over the conformations observed in solution for disaccharide units. Nevertheless, in spite of the fact that a given glycosidic linkage also has the same starting conformation, the influence of the distinct positions in the protein sequence (and of the antennae themselves) over the glycans may still be observed, originating distinct relative abundances of conformations, as noted in α-D-Man*p*-(1→2)-α-D-Man*p* (Fig. 4A–D). This points to the adequacy of the employed protocol in obtaining reliable models for glycoproteins conformational behavior in aqueous solutions.

4. Conclusions

The acquirement of reliable models for the 3D structure of carbohydrates in complex biological systems, such as glycoproteins, represents a challenge for both experimental and theoretical methods. It is a complex picture defined by several possible monomeric units in diverse conformational states, linked in different ways, presenting a variable degree of branching, attached to distinct biological macromolecules and requiring the representation of both spatial and temporal properties in order to be adequately described in solution.

While MD simulations are progressively considered as a potential strategy to obtain a dynamic description of molecules conformational behavior in environments mimicking biological solutions and is successfully applied to the study of proteins, membranes, nucleic acids, and simple carbohydrates, its use to represent glycoproteins and other biological glycoconjugates is much less frequent, mainly due to the lack of parameterization of carbohydrates in its major biological forms. In this context, our group has been dedicated to the conformational representation of glycans of biological significance, in its different levels of complexity, using generally accessible and fast molecular simulation packages.[17,24,25,28–30]

In the current work, we demonstrated that relaxed contour plots obtained from disaccharides employing GROMOS96 43a1 force field[27] added by Löwdin HF/6-31G** atomic charges present similar results to those obtained from MM3 force field. What is most important to mention is that the minimum-energy conformations obtained from such plots represent an adequate starting point for further refinement of disaccharides conformational behavior through MD simulations. This combination of vacuum and explicit solution simulation was capable of offering a reliable platform to describe the conformational ensemble of glycosidic linkages, solely based on the carbohydrate sequence and in the absence of the previous experimental information, reproducing more than 90% of the NMR data of complete glycoproteins. In this context, such an approach may constitute a potential strategy to build accurate models for glycoproteins, as exemplified for COX-1 and COX-2 enzymes, capable of contributing to the understanding of the role of biological glycoconjugates at the atomic level.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (#472174/2007-0), MCT, and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2009.12.018.

References

1. Varki, A. *Glycobiology* **1993**, *3*, 97–130.
2. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
3. Imberty, A.; Gerber, S.; Tran, V.; Pérez, S. *Glycoconjugate J.* **1990**, *7*, 27–54.
4. Brant, D. A. *Pure Appl. Chem.* **1997**, *69*, 1885–1892.
5. Kuttel, M. M.; Naidoo, K. J. *J. Phys. Chem. B* **2005**, *109*, 7468–7474.
6. Xia, J.; Daly, R. P.; Chuang, F.-C.; Parker, L.; Jensen, J. H.; Margulis, C. J. *J. Chem. Theory Comput.* **2007**, *3*, 1620–1628.
7. Xia, J.; Daly, R. P.; Chuang, F.-C.; Parker, L.; Jensen, J. H.; Margulis, C. J. *J. Chem. Theory Comput.* **2007**, *3*, 1629–1643.
8. Xu, Y.; Colletier, J. Ph.; Jiang, H.; Silman, I.; Sussman, J. L.; Weik, M. *Protein Sci.* **2008**, *17*, 601–605.
9. Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–216.
10. Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.* **1997**, *10*, 1249–1261.
11. Sureshan, K. M.; Miyasou, T.; Watanabe, Y. *Carbohydr. Res.* **2004**, *339*, 1551–1555.
12. Crispin, M.; Stuart, D. I.; Jones, E. Y. *Nat. Struct. Mol. Biol.* **2007**, *14*, 354.
13. Berman, H. M.; Henric, K.; Nakamura, H.; Markley, J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 354–355.
14. Cumming, D. A.; Carver, J. P. *Biochemistry* **1987**, *26*, 6664–6676.
15. Wormald, M.; Petrescu, A.-J.; Pao, Y.-L.; Glythero, A.; Elliot, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371–387.
16. Pérez, S.; Mulloy, B. *Curr. Opin. Struct. Biol.* **2005**, *15*, 517–524.
17. Pol-Fachin, L.; Fernandes, C. L.; Verli, H. *Carbohydr. Res.* **2009**, *344*, 491–500.
18. Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D. *Structure* **1994**, *2*, 185–199.
19. Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A. R. N.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269*, 1273–1278.
20. Erbel, P. J. A.; Karimi-Nejad, Y.; van Kuik, J. A.; Boelens, R.; Kamerling, J. P.; Vliegenthart, J. F. G. *Biochemistry* **2000**, *39*, 6012–6021.
21. I.U.P.A.C.-I.U.B. Commission on Biochemical Nomenclature *Pure Appl. Chem.* **1996**, *68*, 1919–2008.
22. Schaftenaar, G.; Noordik, J. H. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123–134.
23. Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr., Sect. D* **2004**, *60*, 1355–1363.
24. Verli, H.; Guimarães, J. A. *Carbohydr. Res.* **2004**, *339*, 281–290.
25. Becker, C. F.; Guimarães, J. A.; Verli, H. *Carbohydr. Res.* **2005**, *340*, 1499–1507.
26. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
27. 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.
28. Verli, H.; Guimarães, J. A. *J. Mol. Graph. Model.* **2005**, *24*, 203–212.
29. Pol-Fachin, L.; Verli, H. *Carbohydr. Res.* **2008**, *343*, 1435–1445.
30. Castro, M. O.; Pomin, V. H.; Santos, L. L.; Vilela-Silva, A.-C. E. S.; Hirohashi, N.; Pol-Fachin, L.; Verli, H.; Mourão, P. A. S. *J. Biol. Chem.* **2009**, *284*, 18790–18800.
31. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
32. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
33. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
34. Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
35. Lütteke, T.; Frank, M.; von der Lieth, C.-W. *Carbohydr. Res.* **2004**, *339*, 1015–1020.
36. Lütteke, T.; Frank, M.; von der Lieth, C.-W. *Nucleic Acids Res.* **2005**, *33*, D242–D246.
37. Lütteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.-W. *Glycobiology* **2006**, *16*, 71R–81R.
38. Mutsaers, J. H. G. M.; van Halbeek, H.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1985**, *147*, 569–574.
39. Nemeth, J. F.; Hochesang, G. P.; Marnett, L. J.; Caprioli, R. M. *Biochemistry* **2001**, *40*, 3109–3116.
40. Lütteke, T.; von der Lieth, C.-W. B. M. C. *Bioinformatics* **2004**, *5*, 69.
41. Furse, K. E.; Pratt, D. A.; Porter, N. A.; Lybrand, T. P. *Biochemistry* **2006**, *45*, 3189–3205.
42. Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K. F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; Kanehisa, M. *Glycobiology* **2006**, *16*, 63R–70R.
43. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8567.
44. Allinger, N. L.; Rahman, M.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293–8307.
45. Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J.-F.; van Eyck, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141–155.
46. Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydr. Res.* **2004**, *339*, 937–948.
47. Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
48. Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; 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., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
49. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
50. Ress, D. A.; Scott, W. E. *J. Chem. Soc., B* **1971**, 469–479.
51. Sathyanarayana, B. K.; Rao, V. S. R. *Biopolymers* **1972**, *11*, 1379–1394.
52. Bluhm, T. L.; Deslandes, Y.; Marchessault, R. H.; Perez, S.; Rinaudo, M. *Carbohydr. Res.* **1982**, *100*, 117–130.
53. Sathyanarayana, B. K.; Stevens, E. S. *J. Biomol. Struct. Dyn.* **1983**, *1*, 947–959.
54. Lipkind, G. M.; Verovsky, V. E.; Kochetkov, N. K. *Carbohydr. Res.* **1984**, *133*, 1–13.
55. Lipkind, G. M.; Shashkov, A. S.; Kochetkov, N. K. *Carbohydr. Res.* **1985**, *141*, 191–197.
56. Virudachalam, R.; Rao, V. S. R. *Biopolymers* **1979**, *18*, 571–589.
57. Shefter, E.; Trueblood, K. N. *Acta Crystallogr.* **1965**, *18*, 1067–1077.
58. Pope, L. *J. Am. Chem. Soc.* **1993**, *115*, 8421–8426.
59. Cumming, D. A.; Carver, J. P. *Biochemistry* **1987**, *26*, 6676–6683.
60. Yoneda, J. D.; Albuquerque, M. G.; Leal, K. Z.; Seidl, P. R.; Wheeler, R. A.; Boesch, S. E.; de Alencastro, R. B.; de Souza, M. C. B. V.; Ferreira, V. F. *Theochem* **2006**, *778*, 97–103.
61. Lütteke, T. *Acta Crystallogr., Sect. D* **2009**, *65*, 156–168.
62. Rollins, T. E.; Smith, W. L. *J. Biol. Chem.* **1980**, *255*, 4872–4875.
63. Otto, J. C.; DeWitt, D. L.; Smith, W. L. *J. Biol. Chem.* **1993**, *268*, 18234–18242.

Supplementary Data

GROMOS96 43a1 performance on the prediction of oligosaccharides conformational ensemble within glycoproteins

Fernandes, C.L.[a], Sachett, L.G. [a], Pol-Fachin, L.[a], and Verli, H.[a,b,*]

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

[b]*Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av. Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil*

Corresponding author. Tel.: +55-51-3308-7770; fax: +55-51-3308-7309; e-mail address: hverli@cbiot.ufrgs.br.

Table S1: Minimum energy geometries for dihedral angles ϕ, ψ and ω, as obtained from relaxed contour plots for a series of disaccharides, refined through 100 ns MD in explicit solvent. Each disaccharide in each pair of dihedrals was submitted to a separate simulation, in a total of 21 simulations (2,100 ns).

Disaccharides	Simulation	ϕ	ψ	ω
β-D-Glc*p*NAc-(1→2)-α-D-Man*p*	A	-60	-150	-
	B	60	90	-
α-D-Man*p*-(1→2)-α-D-Man*p*	A	-180	-90	-
	B	60	-150	-
	C	90	-60	-
α-D-Man*p*-(1→3)-β-D-Man*p*	A	-180	-90	-
	B	60	-120	-
	C	120	90	-
β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc	A	-120	90	-
	B	60	120	-
β-D-Man*p*-(1→4)-β-D-Glc*p*NAc	A	-150	90	-
	B	-60	-60	-
β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc	A	-150	90	-
	B	-90	-60	-
	C	-60	150	-
	D	60	150	-
α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc	A	-150	-150	-60
	B	-90	30	-180
	C	-90	150	60
α-D-Man*p*-(1→6)-β-D-Man*p*	A	60	150	60
	B	90	-120	-180

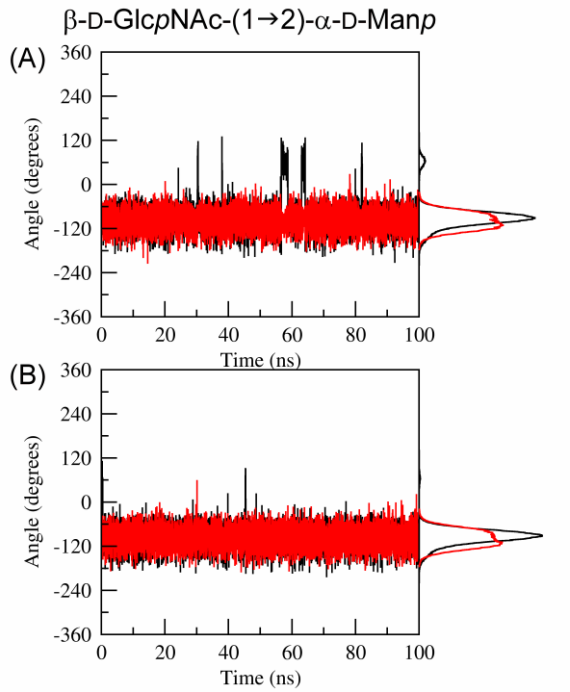


Figure S1: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from disaccharide β-D-Glc*p*NAc-(1→2)-α-D-Man*p* in the MD simulations. Letters (A) and (B) correspond to initial conformations as shown in Table S1.

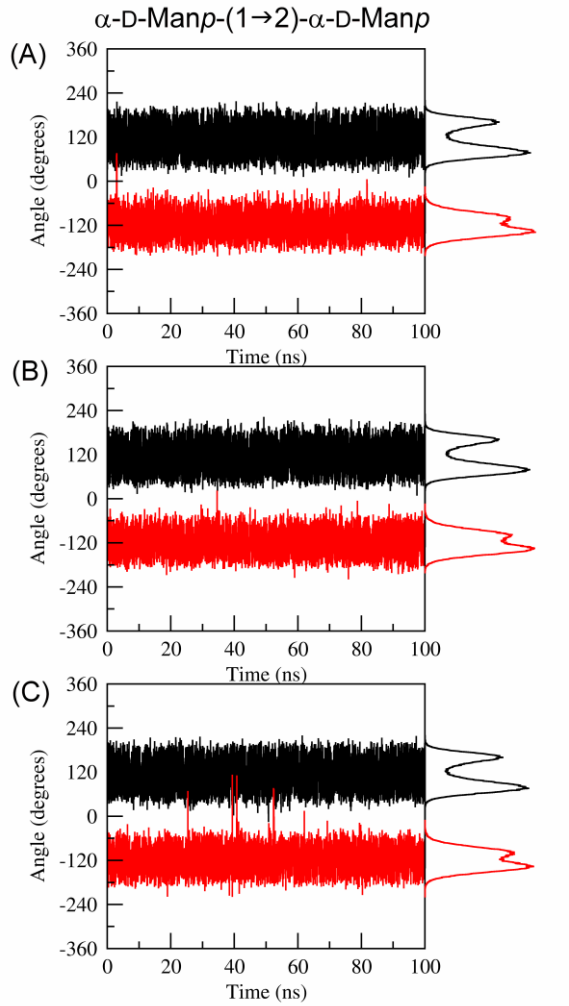


Figure S2: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from in the disaccharide α-D-Man*p*-(1→2)-α-D-Man*p* in MD simulations. Letters (A), (B) and (C) correspond to initial conformations as shown in Table S1.

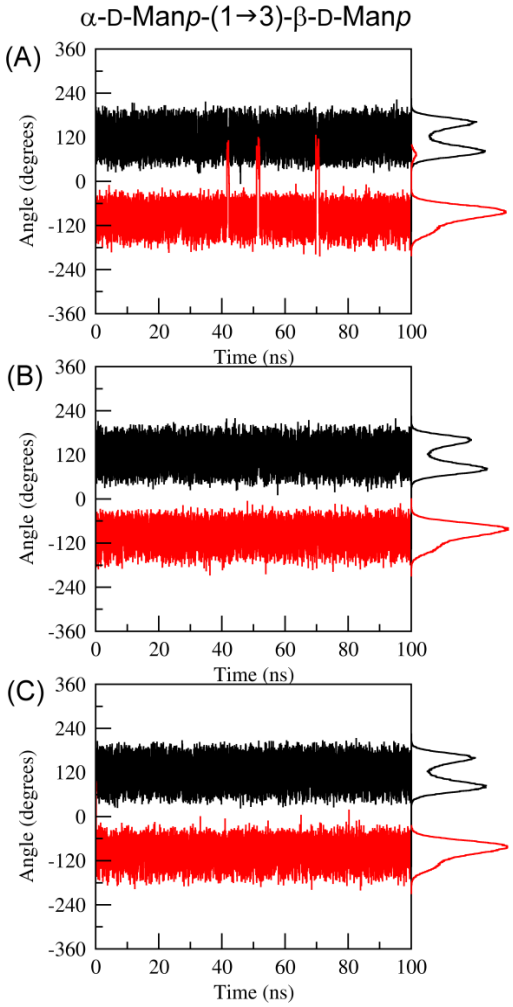


Figure S3: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from in the disaccharide α-D-Man*p*-(1→3)-α-D-Man*p* in MD simulations. Letters (A), (B) and (C) correspond to initial conformations as shown in Table S1.

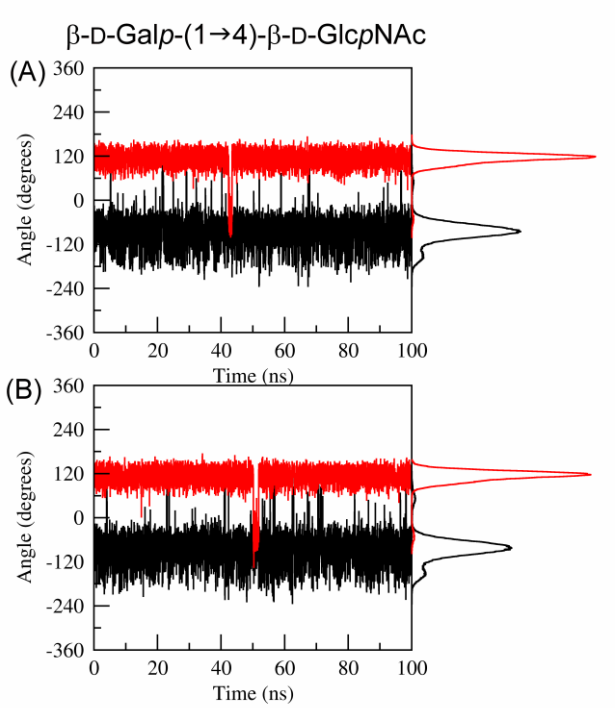


Figure S4: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from in the disaccharide β-D-Gal*p*-(1→4)-β-D-Glc*p*NAc in MD simulations. Letters (A) and (B) correspond to initial conformations as shown in Table S1.

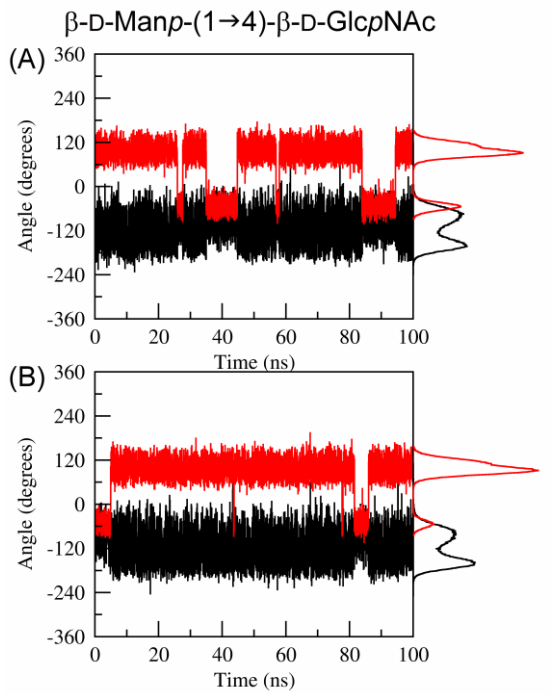


Figure S5: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from in the disaccharide β-D-Man*p*-(1→4)-β-D-Glc*p*NAc in MD simulations. Letters (A) and (B) correspond to initial conformations as shown in Table S1.

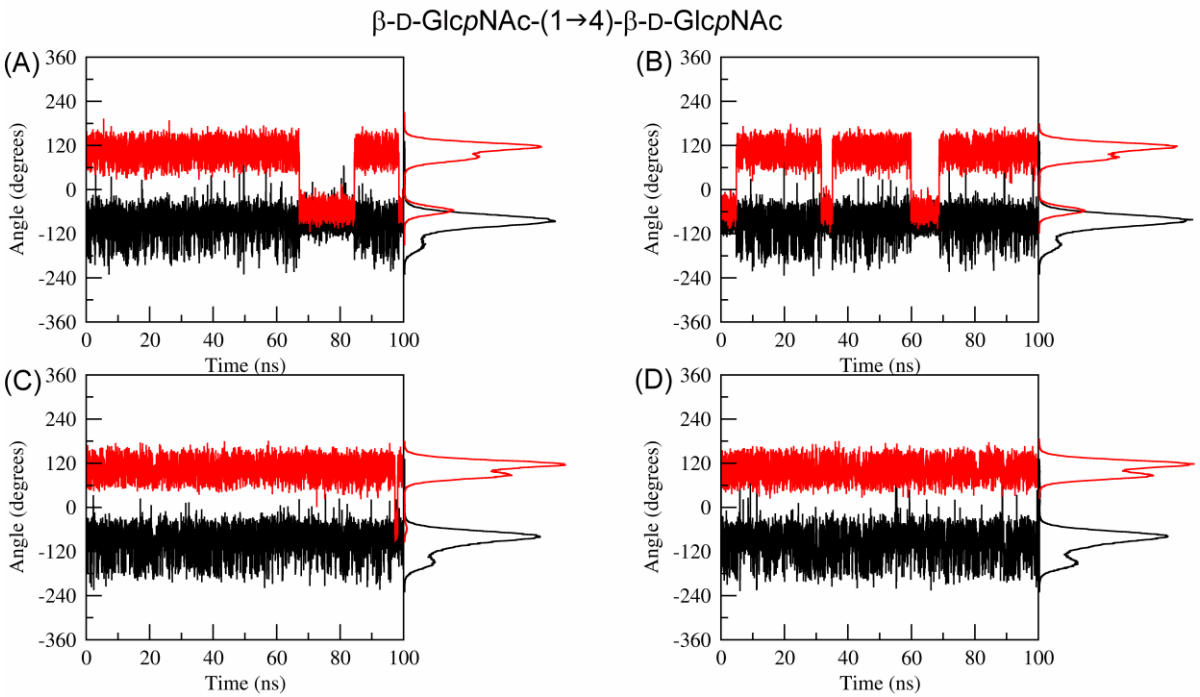


Figure S6: Fluctuation and distribution of ϕ (black) and ψ (red) dihedral angles from in the disaccharide β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc in MD simulations. Letters (A), (B), (C) and (D) correspond to initial conformations as shown in Table S1.

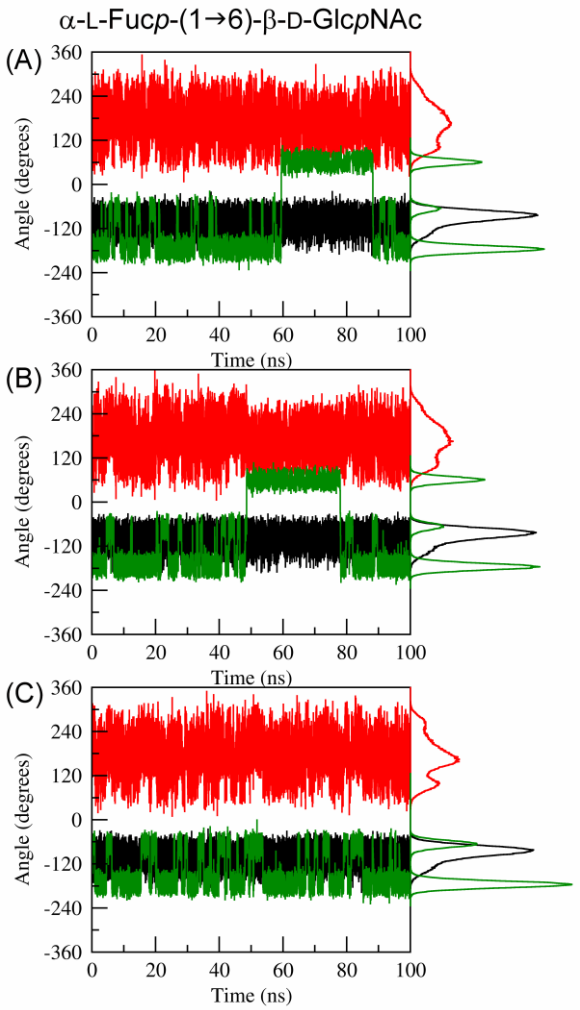


Figure S7: Fluctuation and distribution of ϕ (black), ψ (red) and ω (green) dihedral angles from in the disaccharide α-L-Fuc*p*-(1→6)-β-D-Glc*p*NAc in MD simulations. Letters (A), (B) and (C) correspond to initial conformations as shown in Table S1.

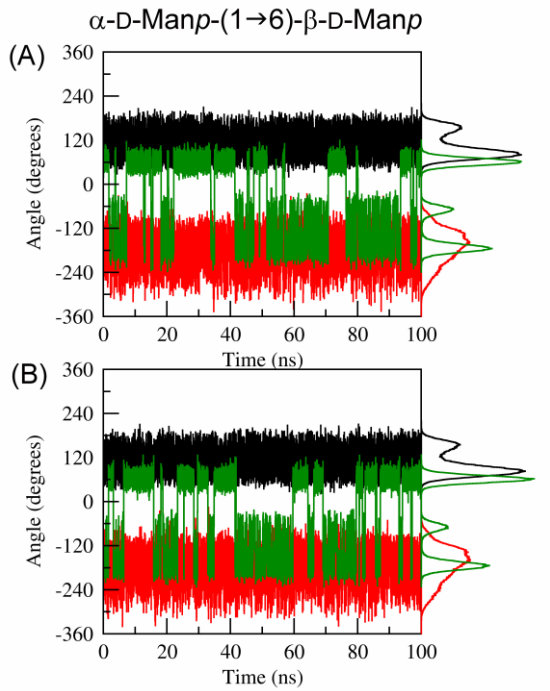


Figure S8: Fluctuation and distribution of ϕ (black), ψ (red) and ω (green) dihedral angles from in the disaccharide α-D-Man*p*-(1→6)-β-D-Man*p* in MD simulations. Letters (A) and (B) correspond to initial conformations as shown in Table S1

Table S2. Glycosylation structures obtained from crystallographic structures available for COX-1. Pdb-care errors found in carbohydrates are shown as red.

PDB	Glycosylation
1PRH 3.5 Å	No glycosylation structure
1EQG 2.6 Å	No glycosylation structure
1EQH 2.7 Å	No glycosylation structure
1HT5 2.7 Å	No glycosylation structure
1HT8 2,7 Å	No glycosylation structure
1PGG 4.5 Å	β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→N)-Asn410
1PGF 4.5 Å	β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→N)-Asn410
1PGE 3.5 Å	β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→N)-Asn410
1PTH 3.4 Å	β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→N)-Asn410
2OYE 2.8 Å	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Man*p*-(1→3)-β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
2OYU 2.7 Å	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1IGX 3.1 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 α-D-Man*p*-(1→3)-β-D-Man*p*-(1→6)-β-D-Man*p*- (1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1IGZ 2.9 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 α-D-Man*p*-(1→3)-β-D-Man*p*-(1→6)-β-D-Man*p*- (1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1FE2 3.0 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Man*p*-(1→3)-β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1DIY 3.0 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 α-D-Man*p*-(1→3)-β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1CQE 3.1 Å	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1Q4G 2.0 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 α-D-Man*p*-(1→6)-α-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144 β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68 β-D-Man*p*-(1→6)-β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144

	β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
2AYL 2.0 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Man*p*-(1→6)-β-D-Man*p*-(1→6)-β-D-Man*p*–(1→4)-α-D-Glc*p*NAc -(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1U67 3.1 Å	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Man*p*-(1→6)-β-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
	α-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn410
1EBV 3.2 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410

Table S3. Glycosylation structures obtained from crystallographic structures available for COX-2.

PDB	Glycosylation
1PXX 2.9 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc -(1→4)-β-D-Glc*p*NAc -(1→4)-β-D-Glc*p*NAc -(1→N)Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
1CVU 2.4 Å	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68
	α-D-Man*p*-(1→6)-α-D-Man*p*-(1→4)-β-D-Glc*p*NAc -(1→4)-β-D-Glc*p*NAc -(1→N)Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
1DDX 3.0 Å	β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
1CX2 3.0 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
3PGH 2.5 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
4COX 2.9 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
5COX 3.0 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410
6COX 2.8 Å	β-D-Glc*p*NAc-(1→N)-Asn68
	β-D-Glc*p*NAc-(1→N)-Asn144
	β-D-Glc*p*NAc-(1→N)-Asn410

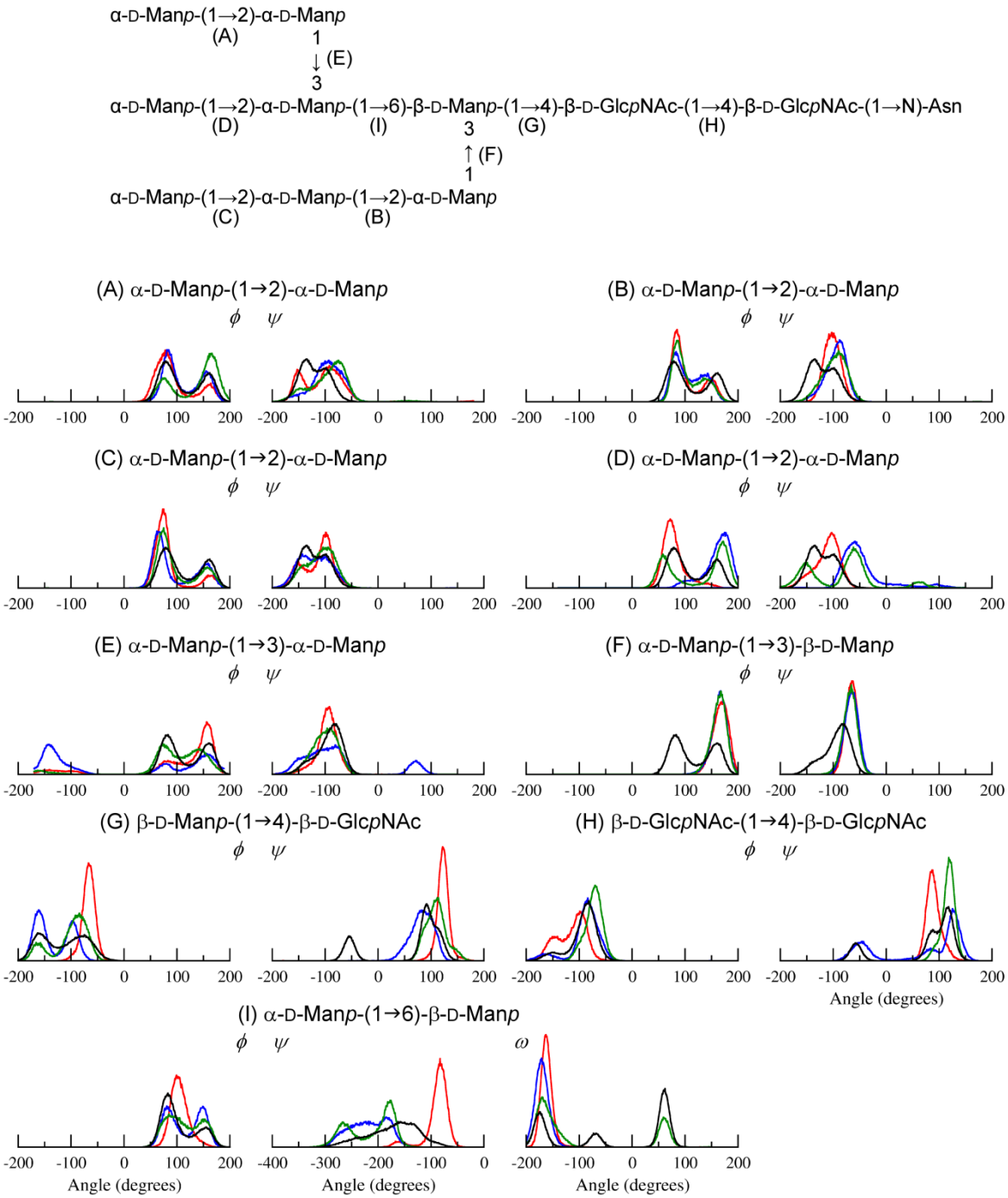


Figure S9: Distribution of the ϕ, ψ and ω dihedral angles, as obtained from MD simulation, from COX-1 glycans at Asn68 (red), Asn144 (blue) and Asn410 (green), together with the conformational patterns associated to the respective disaccharides, isolated in water (black). The A-I letters indicate the relation of a distribution plot to its location on the respective disaccharide unit from the complete glycan.

4.4 Trabalho III

A partir das técnicas de simulação de glicoproteínas desenvolvidas nos dois trabalhos anteriores o presente trabalho tem por objetivo analisar o efeito da glicosilação na estrutura AGP humana. Foram realizadas três simulações: i. AGP não glicosilada; ii. AGP glicosilada com fucoses e iii. AGP glicosilada sem fucoses. A simulação da AGP sem glicosilação monstrou uma grande flexibilidade em relação as formas glicosiladas, demonstrando o papel estrutral da glicosilação na estrutura da proteína. Entretanto nenhum efeito estrutral pode ser relacionado com a presença das fucoses na simulação, sendo necessário um aumento do tempo da simulação para melhor compreender estes eventos.

Como a AGP possui uma cavidade carreadora de fármacos, esta também foi analisada para verificar o efeito da glicosilação sobre ela. Observou-se que a presença das glicanas ajudam na manutenção desta, pois na simulação sem glicosilação a cavidade praticamente desapareceu, enquanto nas formas glicosiladas a cavidade não só se manteve como é capaz de comportar fármacos que são carreados pela AGP.

Adicionalmente foi estudada a interação do tetrassacarídeo SLeX, presente nas simulações com fucoses, com a P-selectina, proteína que se liga a AGP durante a resposta inflamatória. Dos três SLeX presentes na AGP apenas dois tem condições de ligar à P-selectina e o resultado da simulação produziu um encaixe compatível com a interação entre as proteínas. Nota-se um melhor encaixe entre os dois SLeX esdados e isso se deve aos outros glicídeos presentes na vizinhança da ligação. Entretanto mais estudos sobre esta interação são necessários para entender sua base molecular.

Este manuscrito está formatado para ser submetido a revista *Biochemistry.*

Structural glycobiology of human α_1-acid glycoprotein and its implication for drug and P-selectin binding

Fernandes, C.L.[a] and Verli, H.[a,b,*]

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

[b]*Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil*

[*]Corresponding author. Tel.: +55-51-3308-7770; fax: +55-51-3308-7309; e-mail address: hverli@cbiot.ufrgs.br.

Abstract

Human α_1-acid glycoprotein (AGP) is an abundant human plasma glycoprotein, N-glycosylated at residues 15, 38, 54, 75, and 85. While it plays important roles on drug pharmacokinetics, it is also reported to rise up to fivefold in certain inflammatory events. In these processes, the glycan chains attached to Asn54, Asn75 and Asn85 may become fucosylated, originating a sialyl-Lewis X epitope, which in turn binds proteins of the selectin family and, consequently, plays an important role on immunomodulatory processes. While the X-ray structure of unglycosylated AGP was reported, the absence of the glycan chains impairs the obtaining of further insights into its structural biology and, ultimately, into its biological function. In this context, the current work intends to obtain further insights into AGP structural glycobiology and function by building a structural model of its fully glycosylated form, taking into account the different glycoforms observed in biological aqueous solutions. Accordingly, the obtained data points to the absence of a major influence of glycosylation on AGP's secondary structure, in good agreement with its crystallographic structure. However, the glycan chains seems able to interfere with the protein dynamics, mainly at the AGP ligand-binding site, pointing to implications of the AGP carbohydrate moiety on its complexation to drugs and bioactive compounds. Finally, the influence of fucosylation on AGP structure and binding to selectins is explored, indicating that selectin may bind to glycan chains linked to Asn54 and Asn75, and this binding seem to involve others glycans like Asn15. Our data provides insights into the participation of carbohydrates on the AGP inflammatory role.

Keywords: AGP, molecular dynamics, glycoproteins, structural glycobiology.

1. Introduction

The alpha-1 acid glycoprotein (AGP), or orosomucoid, is an abundant human plasma glycoprotein, first described in 1950 (*1, 2*), but with its exact biological role still unclear. Several potential physiological activities were described to this protein, such as drug-binding functions (*3*), inflammatory and immunomodulatory effects (*4*). The AGP is mainly secreted by hepatocytes, and its expression is regulated by several pro-inflammatory cytokines. An important aspect of AGP is that its concentration may rise up to fivefold in inflammatory events, making this protein one of the major members of the acute-phase protein family (*4*).

AGP has a lipocalin fold (*5*) with an extraordinary N-linked carbohydrate content (ca. 45% w/w), becoming one of the most glycosylated known proteins (*4*). Such glycan chains are attached to asparagine residues at positions 15, 38, 54, 75, and 85, each one with a wide variety of possible carbohydrate compositions and degrees of branching (*4, 6-8*). Such variation on AGP carbohydrate composition has been associated to distinct physiological and cellular events (*4*). In chronic inflammatory conditions as septic shock (*9*) and rheumatoid arthritis (*10*) the degree of branching and fucosylation increases in glycan chains attached to Asn54, Asn75 and Asn85. As a result, the α-NeuAc-(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→3)-α-Fuc epitope, called sialyl-Lewis X (SLe^x), is formed, suggesting the participation of AGP in several physiological events, such as inflammation and cell migration (*9, 11, 12*). For example, SLe^x is the binding epitope for the cell adhesion molecules E- and P-selectins, involved in the inflammatory-dependent adhesion of neutrophils, monocytes or platelets. (*13*) Moreover, in tumor metastasis the increased expression of AGP with SLe^x facilitates the binding to endothelial cells via selectins and these contribute to tumor-cell migration to distant tissues and promote metastasis (*14*). Thus the increase in fucosylation and the consequent biosynthesis of the SLe^x epitope on AGP glycans may represent a mechanism for feedback inhibition of granulocyte extravasation in cases of inflammation (*4*).

Additionally. AGP and serum albumin are the two major proteins responsible for drug binding in humans. Understanding the binding mechanism of these proteins is therefore important when proposing new drug candidates (*15*). The physical-chemical properties of AGP favor the binding to basic or neutral compounds of endogenous and exogenous origins, such as tamoxifen, propranolol, heparin, serotonin, platelet activating factor (PAF), melatonin, histamine, and endogenous steroids (cortisol) (*4*). Structurally the AGP crystal without glycosylation has a pocket with 9-12 Å in diameter and exhibits two positively-charged residues, Arg68 and Arg90 (*5*), but offers no information about glycan influence on the pocket.

In this context, the current work intends to contribute in the characterization of the AGP structural glycobiology. Accordingly, atomic level models were obtained, including the AGP complete glycosylation pattern, as observed in humans, as well as some of its distinct glycoforms and the SLe^x epitope. Upon refinement of these models by molecular dynamics (MD) simulations, the obtained data offer new insights into AGP role on inflammatory processes and interaction with P-selectin, as well as into its carbohydrates on the AGP binding to drugs and other bioactive compounds.

2. Computational Methods

2.1. Glycan nomenclature

The IUPAC nomenclature recommendations and symbols (*16*) have been adopted. The relative orientation of a pair of contiguous carbohydrate residues was described, for different types of linkages, by two or three torsional angles at the glycosidic linkage. A (1→X) linkage, where 'X' is '2', '3', '4' or '6', respectively, the *f* and *y* were defined as shown below (in the NeuAc case the linkage was (2→X) and X were '3' or '6'):

f = O5–C1–OX–CX′

y = C1–OX–CX–C(X - 1)

For a (1→6) linkage, the *w* was defined as:

w = O6–C6–C5–C4

2.2. Construction of glycan chains

2.2.1. Carbohydrate trees building and modeling

The five glycans that bind to AGP were selected according to previous works (*6-8*) with a preference to glycans that binds to SLe^x (Figure 1A). Based on previous studies (*17, 18*), the oligosaccharides were analyzed by their disaccharide moieties, composed of N-acetyl-D-glucosamine (GlcNAc), D-mannose (Man), L-fucose (Fuc), D-galactose (Gal) α-D-neuraminic acid (NeuAc) and the major conformational population of f and y angles of these pairs were used to refine the initial structure of glycosylation.

The human AGP were retrieved from PDB ID 3KQ0 (*5*). The N-glycosylation structures (obtained as described above) were linked to their sites (Asn15, Asn38, Asn54, Asn75 and Asn85). In order to understand the effect of fucosylation on the AGP-glycan structure, three systems were constructed, one with AGP without glycosylation, one with the glycosylation described (Figure 1A) and one with a glycosylation equivalent to the latter but without the three fucose residues of glycans Asn54, Asn75 and Asn85 (Figure 1A). All disaccharides that compose glycan trees were simulated in a previous work (*18*) and the topology and charge information were added to the Gromos 43a1 force field (*19*) according to previous results (*20, 21*).

2.2.2. Glycoprotein modeling

The glycoprotein structures were constructed in the Glycoscience portal (*22*) using the GlyProt modeling tool. These initial 3D models had their glycosidic linkage geometries adjusted to the respective disaccharide major conformational state in solution, obtained from MD simulations, as previously reported (*17, 18*) to build a initial model for MD simulations.

2.3. Molecular dynamics simulations

2.3.1. AGP molecular dynamics

The three systems were submitted for MD simulations under explicit aqueous solution using the GROMACS simulation suite (*19*). The protein and glycoproteins were solvated in triclinic boxes using periodic boundary conditions and the SPC water model (*23*). The Lincs method (*24*) 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 using the Particle Mesh Ewald method (*25*). Temperature and pressure were kept constant by coupling disaccharides, ions, and solvent to external temperature and pressure baths, with coupling constants of t = 0.1 and 0.5 ps, (*26*) respectively. The dielectric constant was treated as e = 1, and the reference temperature was adjusted to 310 K. The systems were slowly heated from 50 to 310 K, in steps of 5 ps, each one increasing the reference temperature by 50 K. Each simulation was extended to 0.1 ms.

The characterization of AGP structural glycobiology employed MD simulations of three molecular systems: 1) unglycosylated AGP; 2) fucosylated AGP; 3) glycosylated, non-fucosylated, AGP. The two glycosylation forms were identical except by the three fucoses that form the SLe^x in the Asn54, Asn75 and Asn85 glycosylation positions (Figure 1A). These systems were analyzed taking the crystal structure (*5*) as reference, as well the disaccharide linkages conformational ensemble in solution (*18*), in order to characterize the mutual influence between carbohydrate and protein moieties.

2.3.2. AGP-P-Selectin complex molecular dynamics

To construct the AGP-P-selectin complex the structures employed were the crystallographic model of P-selectin complexed with SLe^x under PDB ID 1G1S (*27*) were the disaccharides angles measure and compare with the angles find in SLe^x AGP simulation. The AGP SLe^x models that satisfied the conformational angles were chosen to a geometrical

docking between SLe^x from P-selectin crystal and AGP simulated. During the simulation only the glycan bond to Asn54 and Asn75 were able to connect to P-selectin due to Asn85 the SLe^x portion present a core position in the glycoprotein, then the models construct were to P-selectin bond through Asn54 and Asn75. The simulations use the same protocol to other simulation (2.3.1) during 50 ns to evaluate the energy and interactions between glycoprotein and P-selectin.

2.4. Cavities measurements

The AGP cavity was measured during the simulation using the dxTuber program (*28*). To evaluate the behavior of cavity during the simulation, 2000 frames (1 nanosecond) were collected from each interval of 10 nanoseconds, totalizing eleven measures for each simulation. For the cavity volume measurements with dxTuber (*28*) dx files from the trajectory were obtained with VMD program (*29*), with the VolMap tool. The protocol employed for dxTuber was its default settings: 3D scanning with a protein density threshold of 0.005 protein atoms/$Å^3$, solvent density threshold of 0.01 atoms/$Å^3$ and minimum diameter of 1 Å.

To compare the cavity volume found during the trajectory, two drugs that bind to AGP had their volume measured with the Chimera program (*30*), aiming to compare these measurements with the cavity volume from the trajectory results.

3. Results and discussion

3.1. Carbohydrate dynamics

The glycan chains were built following a proceeding previously described (*17, 18*), including the use of the main conformational states of disaccharidic units as starting geometries, further refined through MD simulations. Such proceeding was shown as able to

properly reproduce the glycan chains conformational space as determined by NMR techniques, as observed in previous works (*17, 18*).

Using such approach, the AGP glycan chains were built, and the observed conformational behavior characterized (Figure S3 to S5). As a general feature, the glycosydic linkages on AGP populates the main conformational states of the respective disaccharidic unit isolated in solution, as observed for the most linkages. In some cases, however, the minor conformational state of the isolated disaccharide or a conformation different from the one found in disaccharides was the preferred conformation adopted in glycans attached on AGP. Linkages in disagreement with disaccharide simulations were only fifteen (11%), nine from the system without fucose and six from the system with fucose. These differences were observed in specific linkages. In the system with fucose the differences were mainly in the Asn85 glycan. This glycan presents differences in linkages β-GlcNAc-(1→2)-β-Man, β-GlcNAc-(1→4)-β-Man, β-Gal-(1→4)-β-GlcNAc and α-NeuAc-(2→3)-β-Gal. Upon close inspection, some particularities of these linkages became clear: the neuraminic acid involved was the component of the SLe^x portion, and further analysis of this linkage revealed that SLe^x was unable to bind to selectin (Section 3.4). A different conformation was also observed when comparing this glycan to the other fucosylated trees. The glycan branch composed by SLe^x was turned to the protein, and form approximately three H-bonds per frame with AGP (2.9±1.4 per frame). This behavior is different from Asn54 and Asn75-bound glycans in which the SLe^x turned itself to the surrounding solvent, not interacting with the protein. Other observed differences in disaccharides concern the Asn85 glycan are possibly related to the observed variations in the glycan-protein interactions. The other differences in the system with fucose were found in glycan Asn38, linkage β-GlcNAc-(1→4)-β-Man, and glycan Asn15, linkage α-NeuAc-(2→3)-β-Gal. These are the smallest glycans on the protein, and possibly the most rigid as well.

In the systems without fucose the main differences were found in a specific linkage, β-Gal-(1→4)-β-GlcNAc (six differences), two branches of Asn15 and Asn85 glycans and one each in Asn54 and Asn75 glycans. The linkage β-GlcNAc-(1→4)-β-GlcNAc presented differences in Asn15 and Asn54 glycans. The α-Man-(1→3)-β-Man linkage in Asn54 glycan presented a different conformation. These differences, mostly in b-1→4 linkages, have their main conformations located in regions of minor conformational states. According to previous works (*31, 32*), GlcNAc b-1→4 linkages have two or three energy minima.

In total we simulated 133 disaccharides in glycoproteins and 89% of them presented results in accordance with the disaccharides' major conformational states in solution (*17, 18*). Still, the conformational behavior is in good agreement with the geometries expected to occur for each linkage, confirming the quality of the simulations.

In general the flexibility of the carbohydrates in the diantennary glycan linked in the position 38 present the minor flexibility (Figure S6) and these findings agree with previous works (*5*) that related such position to the entrance of the AGP binding cavity. This position has some restraints, and the glycan linked to Asn38 have to be a diantennary tree with no neuraminic acid linked (*4*). Concerning the flexibility of the glycan during the simulation, only in Asn85 was observed a different behavior for the systems with and without fucoses (Figure S6). The fucose system presented a major flexibility than the no-fucose system. In the latter system, the portion SLe^x that was exposed to the protein surface moved to the core, establishing interactions with the surrounding amino acids (see section 3.4 to further details).

3.2. Glycan chain role on protein moiety dynamics

The MD simulation of AGP under both glycosylated and non-glycosylated states supported the obtaining of insights into the influence of the carbohydrate residues on AGP peptidic segment conformation and dynamics. Accordingly, the obtained data indicates a lowering in the amino acids flexibility upon glycosylation of AGP (Figure 2). Such effect

appears to occur mainly at turns and coils (Figure 2-D), whereas some lesser influence was also observed on the protein secondary structure (Figure 2-C). While such influence of glycan chains in glycoproteins was previously observed by our previous works (*17, 18*) and others (*33*), it was also suggested previously for AGP based circular dichoism (*34*). The minor differences in the secondary structure observed in the MD could be an effect of the insufficient time of simulation to describe the unfolding effects. However the analysis of the secondary structure at the final point of all three simulations compare to the crystal structure in general they maintain the same scaffold of crystal (Figure 1-D and Figure S1).

At the residue level, the RMSF analysis demonstrates a region with a major flexibility in the unglycosylated system around the residue 150 (Figure 2-B). This region corresponds to a great loop between two helices in the C-terminus at the surface of the protein. In the glycosylated systems, due to glycosylation trees, this region was no longer exposed to the surrounding solvent and the glycan trees possibly obstruct the loop movements, decreasing its flexibility. Regarding the two glycosylated simulations, both present a similar behavior at the residue level.

3.3. Drug binding site dynamics

AGP and albumin are the most important drug carriers in plasma (*35*) with the difference that albumin prefers acid molecules while AGP binds preferentially basic and neutral molecules, such as warfarin (*36*) and imatinib (*37*). This preference is explained by the presence of neuraminic acids in the surface of the glycoprotein. Thus AGP is an important glycoprotein in studies of pharmacokinetics and efficacy of many therapeutic compounds and consequently in the design of better drug models (*35, 38, 39*).

In order to identify the ligands' mode of binding to AGP, previous docking studies with diazepam and progesterone were performed aiming to understand their transport (*5*). These studies did not include the glycan portion of the glycoprotein, which is essential for this

binding. Therefore, in our studies the AGP cavity was empty but we tried to identify the influence of glycosylation on the recognition of AGP ligands, and the processes involved in the cavity structural maintenance. Volumes of diazepam and progesterone were measured in UCSF Chimera program (*30*), measuring 245.8 $Å^3$ and 307.4 $Å^3$ respectively. These measures aid the evaluation of cavity volume and the protein viability of carrying a ligand.

Initially all cavities have a volume compatible with the crystal structure and during the initial part of simulations the behavior was similar in all conditions. In the process of simulation the cavity measurements indicate that the unglycosylated AGP presents cavity loss, possibly due to the presence of solvent and absence of glycosylation. This indicates that glycosylation contributes primarily to the maintenance of the cavity. Both glycoprotein simulations present a cavity during the entire simulation time and with volumes in agreement with the ligands tested in previous studies (*5*). Moreover, the glycosylation make up a border surrounding the cavity, increasing its volume. The glycosylation also promotes a charged environment, adequate to bind basic ligands. These effects were more cleary observed in the system with fucose.

Analyzing the role of fucoses in the maintenance of cavity, no important structural differences were found in simulations with or without fucoses concerning these residues. This suggests that the role of fucose in the AGP structure may be related to molecular recognition events, principally in the surface of the glycoprotein.

3.4. Structural implications for the potential participation of AGP on inflammatory processes

Lectin was glycoproteins with anti-inflammatory activity and present in acute fase process (*40*). These class of molecules include the selectin, a type of lectin responsible to recognize the glycan SLe^x. The binding of AGP to selectin block this glycoprotein to bind in

leucocites and can be responsible to modulating extravasation into inflamed tissues (*4*). The structure of P-selectin was solved attached to SLe^x (*27*) suggest the conformation of SLe^x necessary to binding to P-selectin. In the crystal (PDB ID 1G1S) exist two molecules of SLe^x binding to P-selectin in presence of ion calcium. This ion provide a hexa coordination of P-selectin and the fucose to binding the SLe^x. The measure of the ϕ and ψ disaccharide dihedral angles of SLe^x, we find ϕ=-75 and ψ=-100 to α-Fuc-(1$\rightarrow$3)-β-GlcNAc; ϕ_1=-70, ϕ_2=-80 and ψ=140 to β-Gal-(1$\rightarrow$4)-β-GlcNAc; ϕ=60 and ψ_1=-110, ψ_2=-120 to α-NeuAc-(2$\rightarrow$3)-β-Gal.

In the simulations the measurement of these disaccharides were performed on SLe^x in order to compare and evaluate the possibility of SLe^x from AGP binding to P-selectin (Figure 5). The results obtained for three SLe^x portions were in accordance with P-selectin from crystallography for glycans bound to Asn54 and Asn75. In the case of Asn85, the ϕ angle of β-Gal-(1$\rightarrow$4)-β-GlcNAc linkages and the ϕ and ψ angles of α-NeuAc-(2$\rightarrow$3)-β-Gal have different conformations from those observed in P-selectin-SLe^x binding (PDB ID 1G1S). In order to observe the binding of P-selectin to AGP, the choice of SLe^x in the same orientation of the crystallographic molecule aided in the process of conformational selection (*41*). This approach took advantage of the functional selectivity of P-selectin binding to SLe^x from the AGP glycan.

The search for SLe^x conformations that were similar to those found in the P-selectin crystal resulted in some conformations compatible with binding in all glycan positions during the simulation. However, in the case of Asn85 glycan, the disaccharide angles differ in most of the frames. Moreover, when we observe the structural position of SLe^x in the Asn85 glycan tree, the localization of SLe^x was not able to bind to P-selectin due to the orientation of SLe^x that was hidden in relation to the surface of AGP. This indicated that the P-selectin and SLe^x connection would make the molecules of AGP and P-selectin overlap. To confirm the proximity of AGP and SLe^x, the interactions between Asn85 SLe^x portion and AGP protein

portion were measured by computing H-bonds during the entire simulations. These measures indicated that each simulation frame had 2.9±1.4 bonds.

Taking this observation into account, the SLe^x present in position Asn54 and Asn75 glycans with conformational orientation similar to the crystallographic model was docked to P-selectin in presence of calcium ion. In both positions the results fit the SLe^x present in the glycan of AGP molecule in the binding of P-selectin with distance among calcium ion, fucose and P-selectin favorable to coordination, demonstrating the potential to form these complex (Figure. 5).

The interaction of P-selectin and SLe^x was based on crystallographic data (*27*), by overlapping of SLe^x portion in crystallographic model and MD simulation o and calcium coordination.. For each SLe^x in glycan Asn54 and Asn75 we constructed one initial model for AGP-P-selectin (Figure 5). The P-selectin-AGP overlap hindered the modeling of the Asn85 glycan.

The MD simulation of both models demonstrated an interaction between AGP and P-selectin. The energy between the specific binding glycan and P-selectin was measured during the simulation in order to verify if the molecules were bonded (Figure 6). To analyze the total energy between glycan and P-selectin, measurements were performed from 10 ns of simulation onwards, in order to obtain results after system equilibration. The energy in the complex bond by Asn54 glycan were more stable than the one formed by Asn75 glycan, with energy of -526 ± 69 kJ/mol for the former and -472 ± 108 kJ/mol for the latter.

P-selectin forms a complex with AGP by binding specifically to a glycan composed by the SLe^x portion. These complexes, however, demonstrate differences regarding the participation of the other glycans in the process. In order to understand these processes, the formation of H-bonds between P-selectin and all glycans was measured. In the complex formed by P-seletin and AGP binding to Asn54 glycan this glycan forms 7.9 ± 2.6 H-bonds

per frame and specifically 4.7 ± 1.6 of these H-bonds involve the SLe^x portion. Analyzing the other glycans we identified that Asn15 interacted with P-selectin forming 6.3 ± 2.5 H-bonds per frame. These bonds seem to stabilize the interaction of the complex.

The complex formed by P-selectin and Asn75-bound glycan formed 6.4 ± 2.9 H-bonds and 2.1 ± 1.6 with the SLe^x portion. Regarding other glycans that interact with this complex only the one bound to Asn38 showed significant interactions, with 1.8 ± 1.4 H-bonds per frame. The differences in the mode of binding observed for the two complexes may explain the energetic differences in the complex formed by Asn54-bound glycan when compared to the one formed with the Asn75-bound glycan. The glycan from Asn54 may owe its enhanced stability to its interaction with the Asn15-bound glycan. The Asn38-bound glycan has only minor interactions in the complex formed by Asn75-bound glycan.

These MD simulations experiments provided a basis for the molecular recognizing of AGP by P-selectin. The simulations of AGP and P-selectin binding to either Asn54 or Asn75 glycans were performed and the interaction energy between the AGP glycan and P-selectin were measured. These energy measurements indicated that both molecules interact during the analyzed period for both interaction cases. Moreover, the interaction with Asn54-bound glycan is more stable than the one with Asn75-bound glycan, mainly due to stabilization involving the glycan bound to Asn15-glycan.

4. Conclusions

In this work we were able to obtain models of AGP in two glycosylated forms employing its crystallographic structure and the sequences of glycosylation of the five glycosylation sites obtain from mass spectroscopy. Most important, such models accounts for dynamical features of AGP in aqueous solution as derived from MD simulations.

Accordingly, the protein dynamics demonstrated a major flexibility of the unglycosylated structure when compared to its glycosylated forms, pointing to a glycosylation-induced stabilization of the protein.

The different glycosylation forms do not interfere with the flexibility of glycans in general. The principal influence of glycosylation was related to cavity preservation. Therefore to better understand the role of fucosylation in the flexibility simulations a large time scale would be necessary.

The binding of AGP to P-selectin was shown to be possible in the positions Asn54 Asn75 and these findings suggest that AGP has conformational requirements to recognize P-selectin in inflammatory events due to the possibility to coordinate calcium ion between fucose and P-selectin. However the position Asn85 are not able to dock to P-selectin due to de SLe^x in this position being in the core of the glycoprotein and the disaccharide dihedrals are in disagreement with crystallographic data for P-selectin. The MD initial models are in accordance with P-selectin crystal distances for P-selectin and SLe^x (*27*), with this interaction depend ending on fucose and calcium ion. In this work was observed the interaction of AGP and P-selectin through the different glycan trees bound to Asn54 and Asn75. It is known that P-selectin interacts with different SLe^x-containing glycosylation trees (*4*). During the simulation time both the studied glycan trees were able to keep the proteins bound. Considering the overwhelming variation observed for AGP glycoforms, it is interesting to point out the importance of the glycan structures surrounding the SLe^x-containing tree, since these non-binding glycans seem to take part in the complex scenario of AGP-P-selectin interaction, as observed in the present work.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, MCT, by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil.

References:

1. Schmid, K. (1950) Preparation and properties of an acid glycoprotein prepared from human plasma, *Journal of the american chemical society 72*, 2816-2816.
2. Schmid, K. (1953) Preparation and Properties of Serum and Plasma Proteins. XXM. Separation from Human Plasma of Polysaccharides, Peptides and Proteins of Low Molecular Weight. Crystallization of an Acid Glycoprotein, *Journal of the american chemical society 75*, 60-68.
3. Wasan, K. M., Brocks, D. R., Lee, S. D., Sachs-Barrable, K., and Thornton, S. J. (2008) Impact of lipoproteins on the biological activity and disposition of hydrophobic drugs: implications for drug discovery, *Nature Reviews Drug Discovery 7*, 84-99.
4. Fournier, T., Medjoubi-N, N., and Porquet, D. (2000) Alpha-1-acid glycoprotein, *Biochimica et Biophysica Acta (BBA) - General Subjects 1482*, 157-171.
5. Schonfeld, D., Ravelli, R., Mueller, U., and Skerra, A. (2008) The 1.8-Å Crystal Structure of α1-Acid Glycoprotein (Orosomucoid) Solved by UV RIP Reveals the Broad Drug-Binding Activity of This Human Plasma Lipocalin, *Journal of Molecular Biology 384*, 393-405.
6. Nakano, M., Kakehi, K., Tsai, M.-H., and Lee, Y. C. (2004) Detailed structural features of glycan chains derived from a1-acid glycoproteins of several different animals: the presence of hypersialylated, O-acetylated sialic acids but not disialyl residues, *Glycobiology 14*, 431-441.
7. Sei, K., Nakano, M., Kinoshita, M., Masuko, T., and Kakehi, K. (2002) Collection of a1-acid glycoprotein molecular species by capillary electrophoresis and the analysis of their molecular masses and carbohydrate chains Basic studies on the analysis of glycoprotein glycoforms, *Journal of Chromatography A 958*, 273-271.
8. Nagy, K., Vekey, K., Imre, T., Ludanyi, K., Barrow, M. P., and Derrick, P. J. (2004) Electrospray ionization fourier transform ion cyclotron resonance mass spectrometry of human alpha-1-acid glycoprotein, *Analytical Chemistry 76*, 4998-5005.
9. Linden, E. C. M. B.-V. D., Ommen, E. C. R. V., and Dijk, W. V. (1996) Glycosylation of alpha-acid glycoprotein in septic shock: changes in degree of branching and in expression of sialyl Lewis" groups, *Glycoconjugate Journal 13*, 27-31.
10. Olewicz-Gawlik, A., Korczowska-Lacka, I., Lacki, J. K., Klama, K., and Hrycaj, P. (2007) Fucosylation of serum α1-acid glycoprotein in rheumatoid arthritis patients treated with infliximab, *Clinical Rheumatology 26*, 1679-1684.
11. Graaf, T. W. D., Stelt, M. E. V. d., Anbergen, M. G., and Dijk, W. v. (1993) Inflammation-induced Expression of Sialyl Lewis X-containlng Glycan Structures on alpha-1-Acid Glycoprotein (Orosomucoid) in Human Sera, *J. Exp. Med. 177*, 657-666.
12. Levander, L., Gunnarsson, P., Grenegård, M., Rydén, I., and Påhlsson, P. (2009) Effects of α1-acid Glycoprotein Fucosylation on its Ca2+Mobilizing Capacity in Neutrophils, *Scandinavian Journal of Immunology 69*, 412-420.

13. Lasky, L. A. (1992) Selectins: interpreters of cell-specific carbohydrate information during inflammation, *Science 258*, 964-969.
14. Dube, D. H., and Bertozzi, C. R. (2005) Glycans in cancer and inflammation--potential for therapeutics and diagnostics, *Nat Rev Drug Discov 4*, 477-488.
15. Yasgar, A., Furdas, S. D., Maloney, D. J., Jadhav, A., Jung, M., and Simeonov, A. (2012) High-throughput 1,536-well fluorescence polarization assays for alpha(1)-acid glycoprotein and human serum albumin binding, *PLoS One 7*, e45594.
16. (1996) IUPAC-IUB Commission on Biochemical Nomenclature, *Pure Appl. Chem 68*, 1919-2008.
17. Pol-Fachin, L., Fernandes, C. L., and Verli, H. (2009) GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations, *Carbohydrate Research 344*, 491-500.
18. Fernandes, C. L., Sachett, L. G., Pol-Fachin, L., and Verli, H. (2010) GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins, *Carbohydrate Research 345*, 663-671.
19. Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free, *Journal of Computational Chemistry 26*, 1701-1718.
20. Verli, H., and Guimaraes, J. A. (2004) Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution *Carbohydrate Research 339*, 281-290.
21. Becker, C. F., Guimaraes, J. A., and Verli, H. (2005) Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution, *Carbohydrate Research 340*, 1499-1507.
22. Lutteke, T., Frank, M., and Lieth, C.-W. v. d. (2004) Data mining the protein data bank: automatic detection and assignment of carbohydrate structures, *Carbohydrate Research 339*, 1015-1020.
23. Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The missing term in effective pair potentials, *Journal of Physical Chemistry 91*, 6269-6271.
24. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations, *Journal of Computational Chemistry 18*, 1463-1472.
25. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N⊕log(N) method for Ewald sums in large systems, *Journal of Chemical Physics 98*, 10089-10093.
26. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. v., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics 81*, 3684-3690.
27. Somers, W. S., Tang, J., Shaw, G. D., and Camphausen, R. T. (2000) Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLeX and PSGL-1, *Cell 103*, 467-479.
28. Raunest, M., and Kandt, C. (2011) dxTuber: detecting protein cavities, tunnels and clefts based on protein and solvent dynamics, *J Mol Graph Model 29*, 895-905.
29. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD - Visual Molecular Dynamics, *Journal of Molecular graphics 14*, 33-38.
30. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.

31. Grigory M . Lipkind, V. E. V., Nikolay K . Kochetkov. (1984) Conformational states of cellobiose and maltose in solution: a comparison of calculated and experimental data, *Carbohydrate Research 133*, 1-13.
32. Grigory M. Lipkind, A. S. S., Nikolay K. Kochetkov. (1985) Nuclear overhauser effect and conformational states of cellobiose in aqueous solution, *Carbohydrate Research 141*, 191-197.
33. Woods, R. J., Pathiaseril, A., Wormald, M. R., Edge, C. J., and Dwek, R. A. (1998) The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule, *European Journal of Biochemistry 258*, 372-386.
34. Ceciliani, F., and Pocacqua, V. (2007) The Acute Phase Protein alpha1-Acid Glycoprotein: A Model for Altered Glycosylation During Diseases, *Current Protein and Peptide Science 8*, 91-108.
35. Trainor, G. L. (2007) The importance of plasma protein binding in drug discovery, *Expert Opin. Drug Discov. 2*, 51-64.
36. Otagiri, M., Maruyama, T., Imai, T., Suenaga, A., and Imamura, Y. (1987) A comparative study of the interaction of warfarin with human alpha 1-acid glycoprotein and human albumin, *J. Pharm. Pharmacol. 39*, 416-420.
37. Fitos, I., Visy, J., Zsila, F., Mady, G., and Simonyi, M. (2006) Selective binding of imatinib to the genetic variants of human α1-acid glycoprotein, *Biochim. Biophys. Acta 1760*, 1704-1712.
38. Parikh, H. H., McElwain, K., Balasubramanian, V., Leung, W., Wong, D., Morris, M. E., and Ramanathan, M. (2000) A rapid spectrofluorimetric technique for determining drug–serum protein binding suitable for high-throughput screening, *Pharm. Res. 17*, 632-637.
39. Watanabe, K., Ishima, Y., Akaike, T., Sawa, T., Kuroda, T., Ogawa, W., Watanabe, H., Suenaga, A., Kai, T., Otagiri, M., and Maruyama, T. (2013) S-nitrosated alpha-1-acid glycoprotein kills drug-resistant bacteria and aids survival in sepsis, *Faseb J 27*, 391-398.
40. Alencar, N. M. N., Oliveira, R. S. B., Figueiredo, J. G., Cavalcante, I. J. M., Matos, M. P. V., Cunha, F. Q., Nunes, J. V. S., Bomfim, L. R., and Ramos, M. V. (2009) An anti-inflammatory lectin from Luetzelburgia auriculata seeds inhibits adhesion and rolling of leukocytes and modulates histamine and PGE2 action in acute inflammation models, *Inflammation Research 59*, 245-254.
41. Vaidehi, N., and Kenakin, T. (2010) The role of conformational ensembles of seven transmembrane receptors in functional selectivity, *Current opinion in pharmacology 10*, 775-781.

Captions to figures:

Figure 1: (A) Schematic representation of human AGP glycosylation. The saccharides involved were b-N-acetyl-D-glucosamine (GlcNAc), a/b-D-mannose (Man), a-L-fucose (Fuc), b-D-galactose (Gal) and a-N-acetylneuraminic acid (NeuAc). Snapshots of MD simulations for human AGP glycosylation structures (B) with fucoses and (C) without fucoses and (D) fit of final dynamics structures unglycosylated (green), with fucoses (blue) and without fucoses (red).

Figure 2: All-atoms protein RMSD (A) and all-atoms protein RMSF (B), RMSD of secondary structure (C) and turns and coils (D).

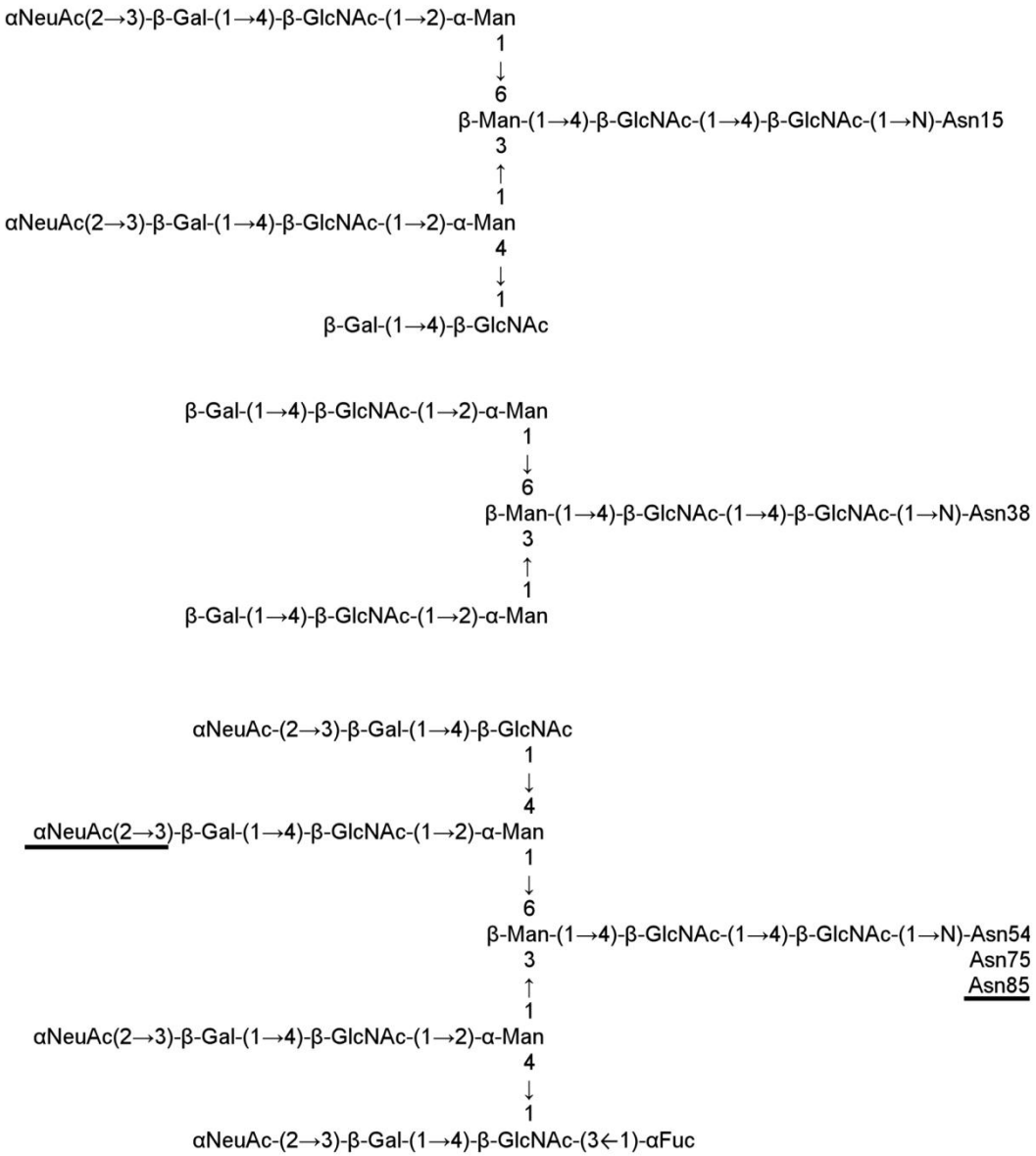
Figure 3: AGP volume cavities found by dxTuber, demonstrating the loss of cavity in the unglycosylated form and the volume of glycosylated-form cavity, which is compatible with the binding of AGP ligands as previously studied (*5*), namely diazepam (245.8 $Å^3$) and progesterone (307.4 $Å^3$).

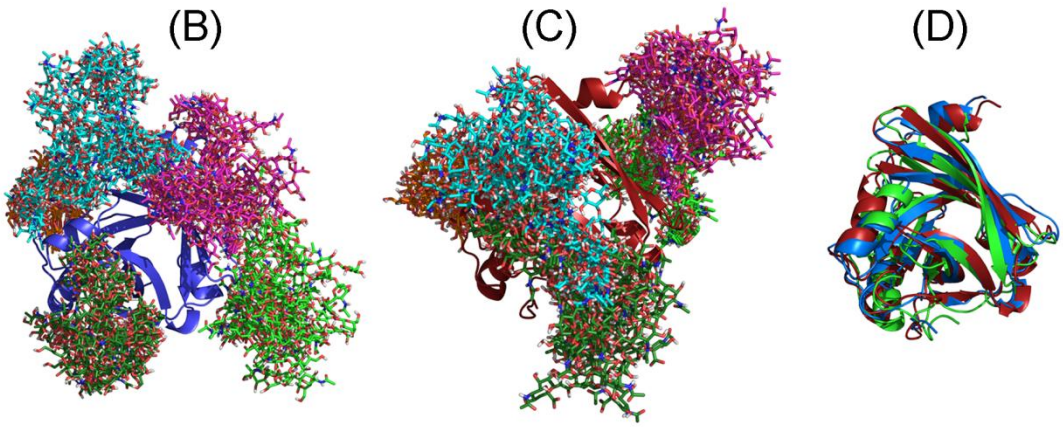
Figure 4: Analysis of the relative abundance of the bioactive conformation of SLe^x in the simulation, as defined in its crystallographic complex with P-selectin (PDB ID 1G1S). The disaccharide dihedrals that are present in the crystal are represented on each fucosylated AGP glycan tree (dotted line in the graphics). The disaccharides that compose each SLe^x are colored as black when attached to Asn54, as red when attached to Asn75, and as green when attached to Asn85.

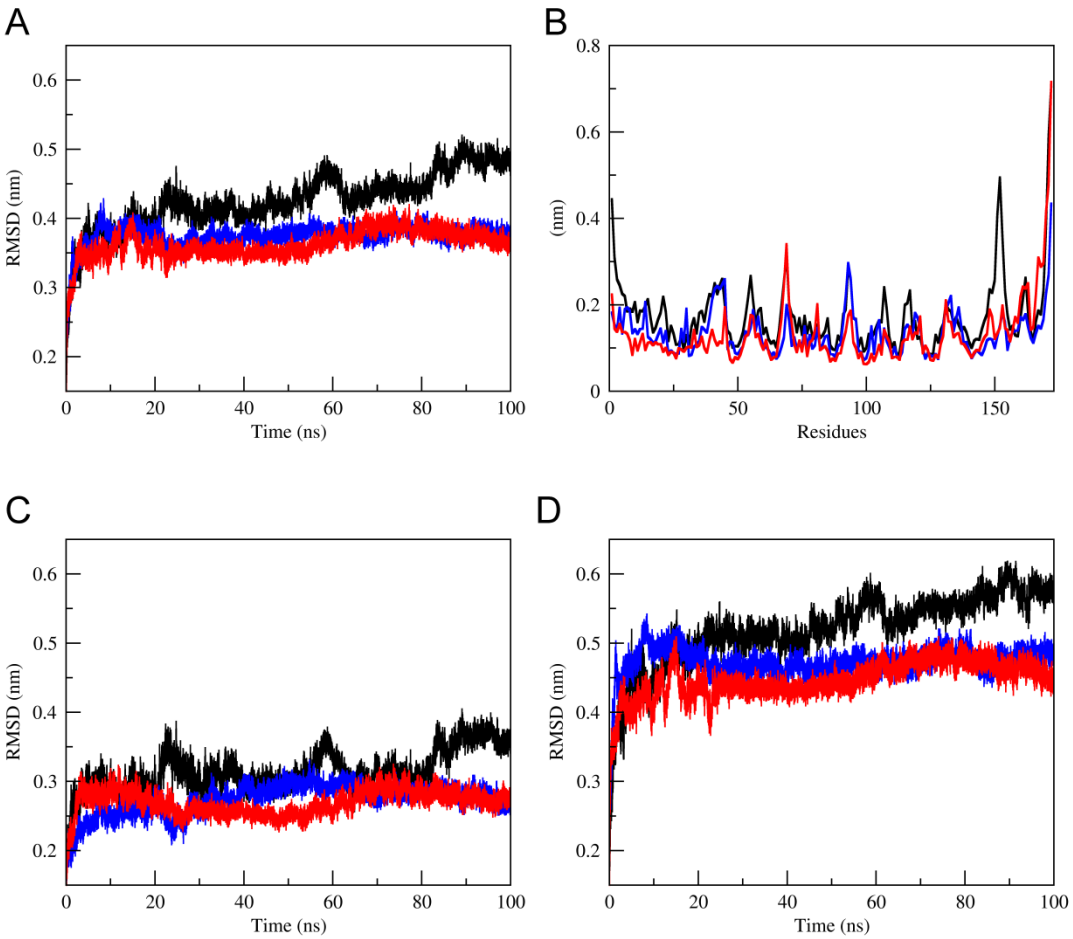
Figure 5: Representation of AGP-P-selectin complex formed in presence of calcium (blue). P-selectin was linked to Asn54-bound glycan (green) and Asn75-bound glycan (brown). The AGP proteins in both complexes were fit and the other glycan trees not involved in the initial interactions were hidden to facilitate visualization of the complexes.

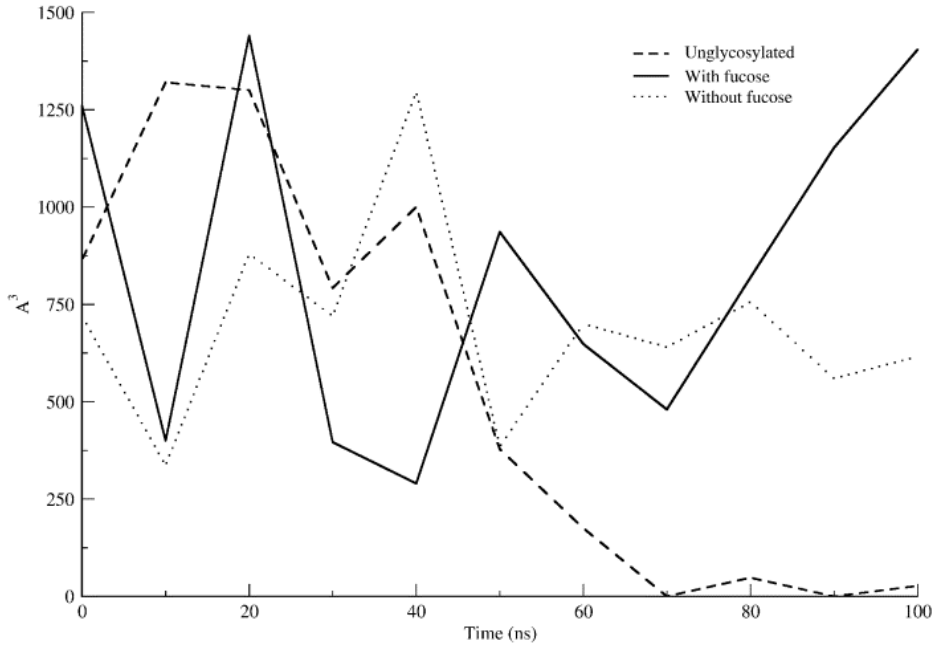
Figure 6: Total energy between the P-selectin and the glycan that comprises the SLe^x.

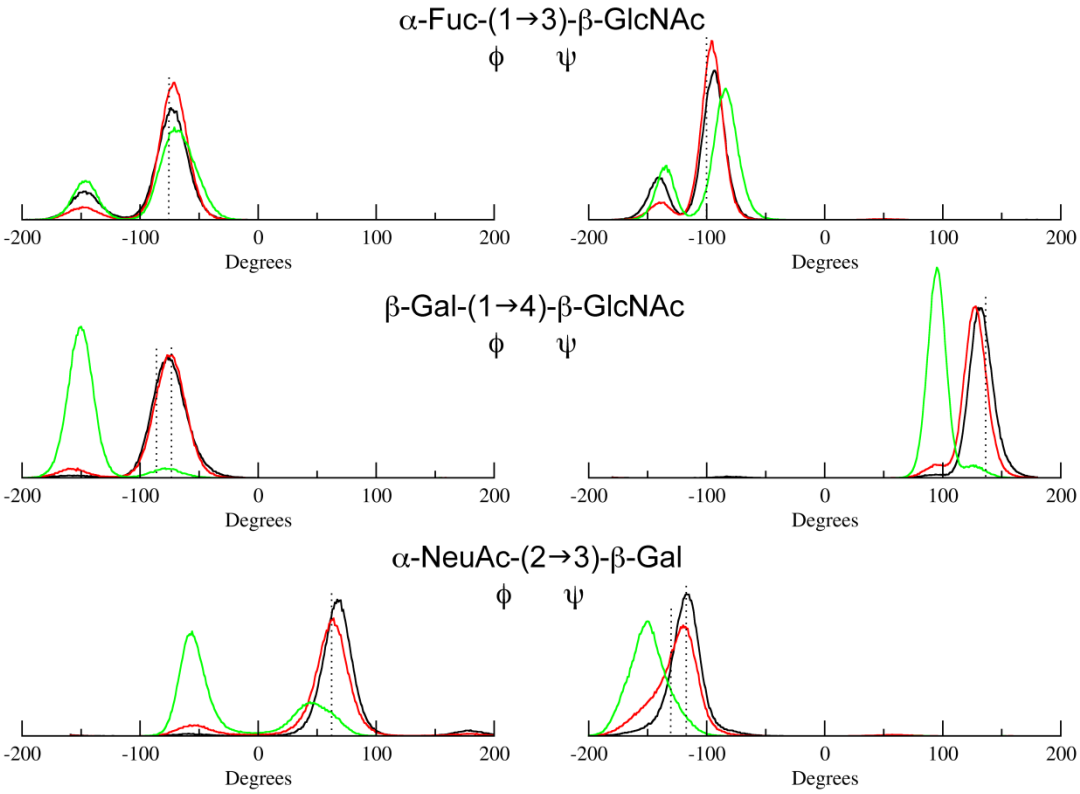
(A)

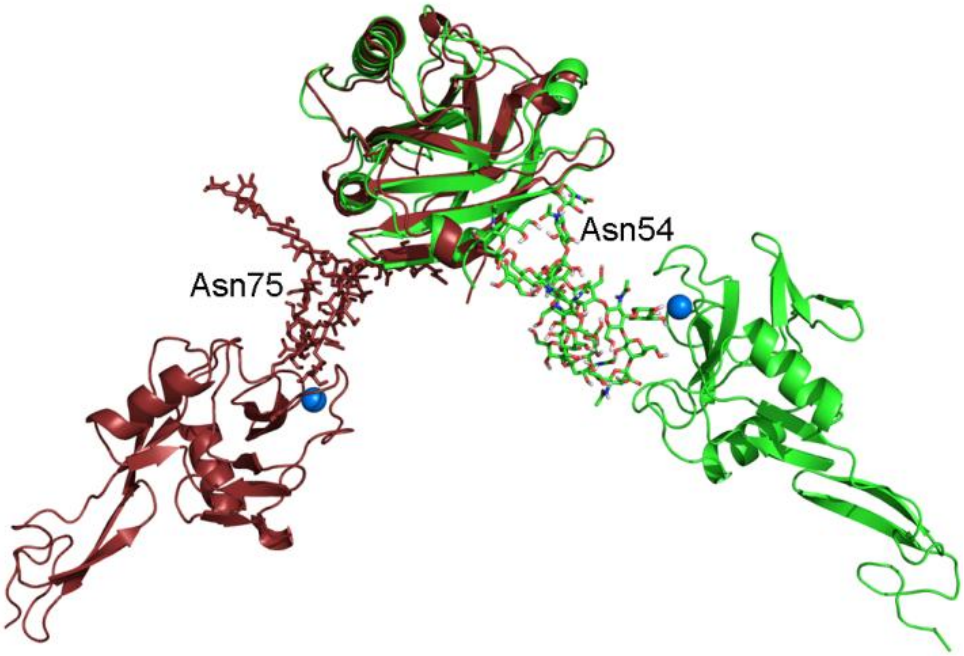

αNeuAc(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
1
↓
6
β-Man-(1→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N)-Asn15
3
↑
1
αNeuAc(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
4
↓
1
β-Gal-(1→4)-β-GlcNAc
β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
1
↓
6
β-Man-(1→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N)-Asn38
3
↑
1
β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
αNeuAc-(2→3)-β-Gal-(1→4)-β-GlcNAc
1
↓
4
αNeuAc(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
1
↓
6
β-Man-(1→4)-β-GlcNAc-(1→4)-β-GlcNAc-(1→N)-Asn54
3
Asn75
↑
Asn85
1
αNeuAc(2→3)-β-Gal-(1→4)-β-GlcNAc-(1→2)-α-Man
4
↓
1
αNeuAc-(2→3)-β-Gal-(1→4)-β-GlcNAc-(3←1)-αFuc

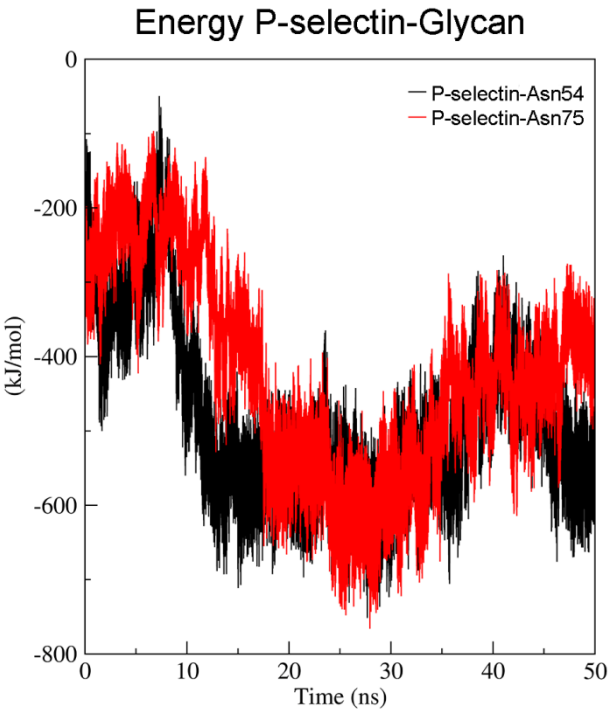

(B)
(C)
(D)


A
RMSD (nm)
Time (ns)
B
(nm)
Residues
C
RMSD (nm)
Time (ns)
D
RMSD (nm)
Time (ns)


Unglycosylated
With fucose
Without fucose
A^3
Time (ns)
0
250
500
750
1000
1250
1500
0
20
40
60
80
100


α-Fuc-(1→3)-β-GlcNAc
ϕ
ψ
-200
-100
0
100
200
Degrees
β-Gal-(1→4)-β-GlcNAc
ϕ
ψ
-200
-100
0
100
200
Degrees
α-NeuAc-(2→3)-β-Gal
ϕ
ψ
-200
-100
0
100
200
Degrees


Asn54
Asn75


Energy P-selectin-Glycan
P-selectin-Asn54
P-selectin-Asn75
0
-200
-400
-600
-800
(kJ/mol)
0
10
20
30
40
50
Time (ns)

Supplementary Data

Structural glycobiology of human α_1-acid glycoprotein and its implication for drug and P-selectin binding

Fernandes, C.L.[a] and Verli, H.[a,b,*]

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

[b]*Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil*

Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil

[*]Corresponding author. Tel.: +55-51-3308-7770; fax: +55-51-3308-7309; e-mail address: hverli@cbiot.ufrgs.br.

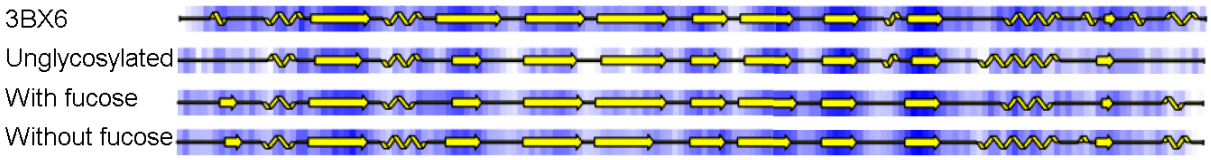


Figure S1: Secondary structure contends of crystal structure of AGP (3KQ0) and the simulated systems.

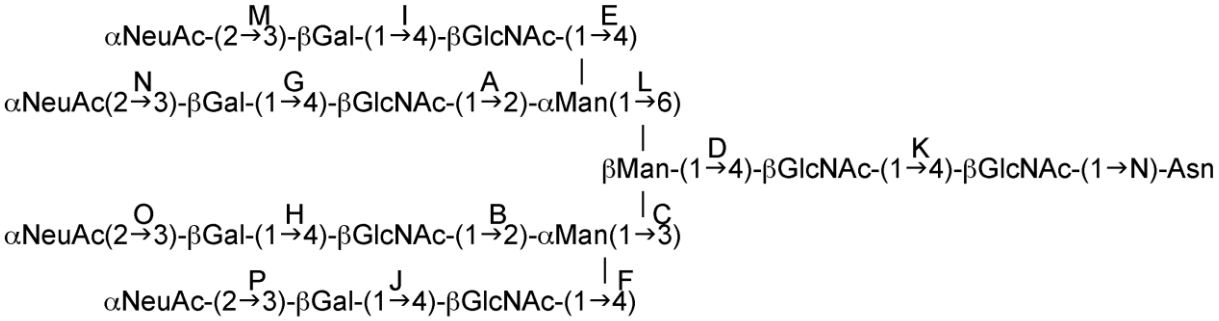


Figure S2: Glycosylation pattern of AGP. Letters up to glycosylation correspond to the dihedrals analyzed in figures S3 to S5 to systems with and without fucose.

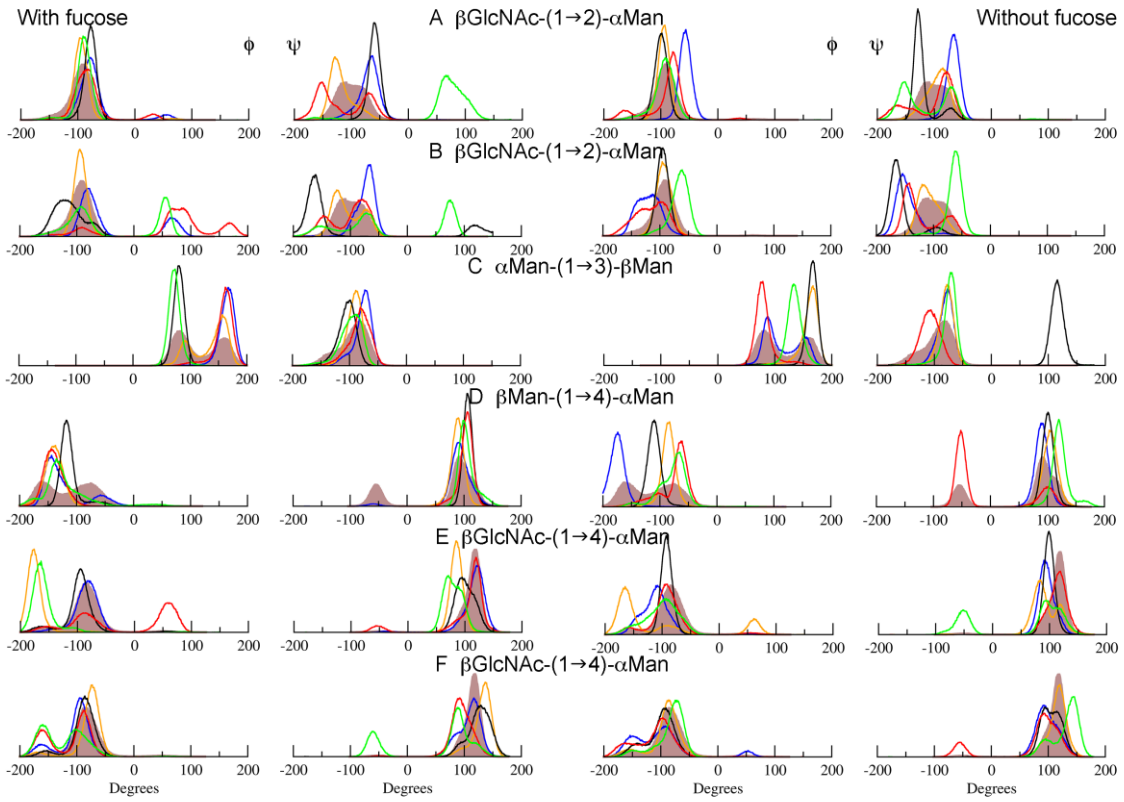


Figure S3: AGP with fucose distribution of the *f*, *y* dihedral angles from glycans at Asn15 (blue), Asn38 (orange), Asn54 (black), Asn75 (red) and Asn85 (green). The coloured curve (brown) indicates the most prevalent conformation of disaccharides in solution.

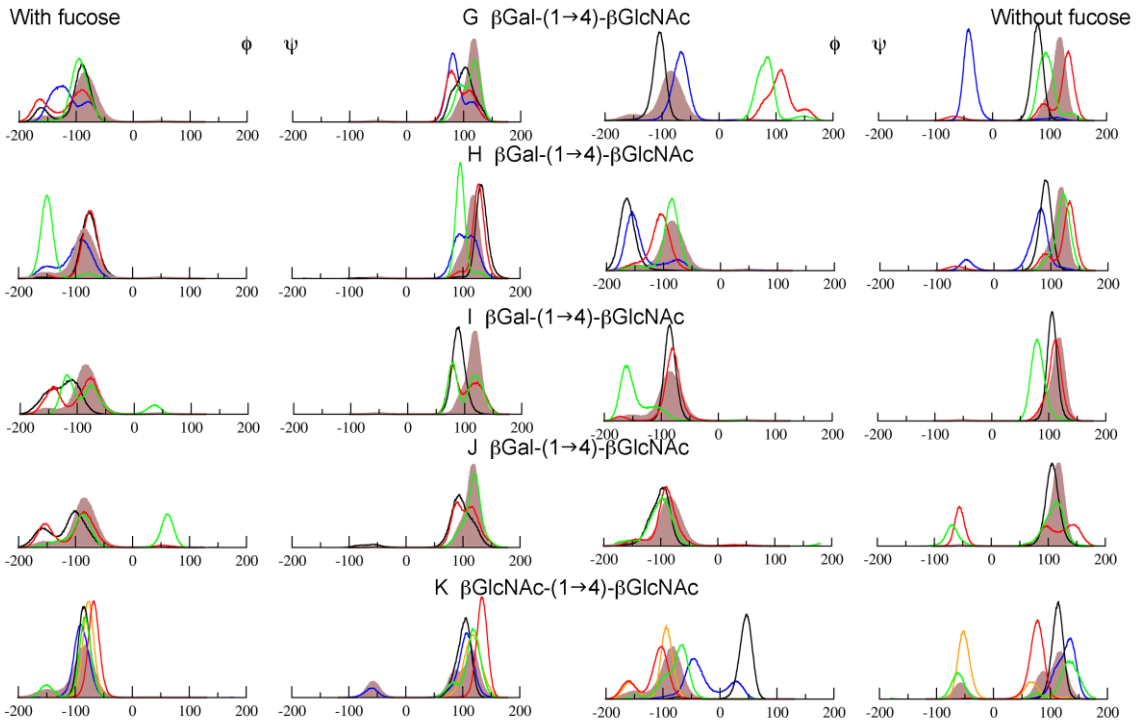


Figure S4: AGP with fucose distribution of the *f* and *y* dihedral angles from glycans at Asn15 (blue), Asn38 (orange), Asn54 (black), Asn75 (red) and Asn85 (green). The coloured curve (brown) indicates the most prevalent conformation of disaccharides in solution.

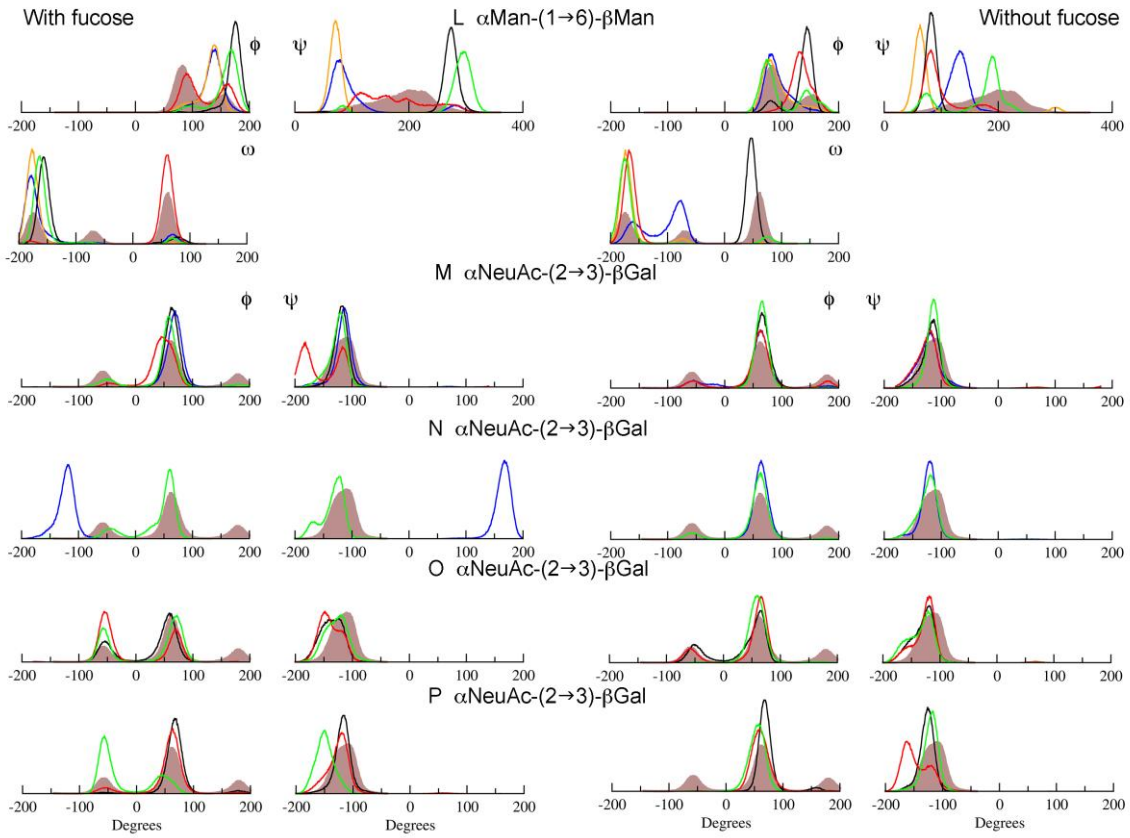


Figure S5: AGP with fucose distribution of the *f* and *y* and *w* dihedral angles from glycans at Asn15 (blue), Asn54 (black), Asn75 (red) and Asn85 (green). The coloured curve (brown) indicates the most prevalent conformation of disaccharides in solution.

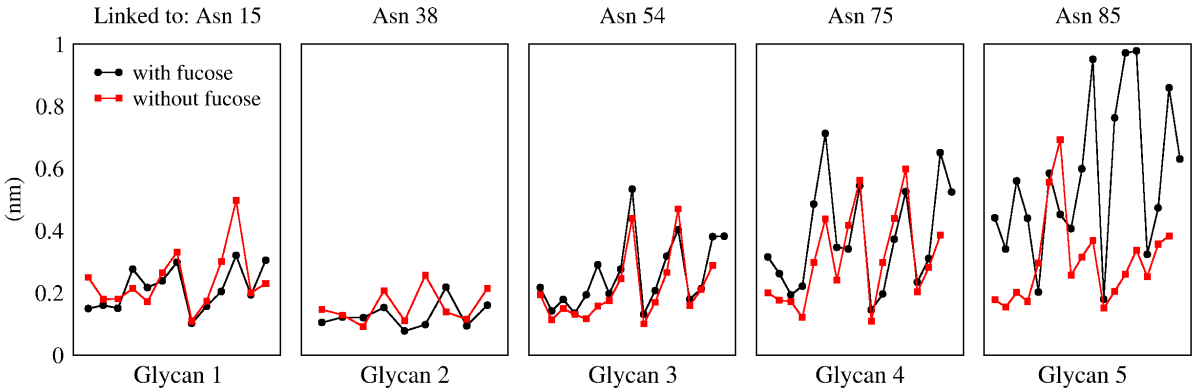


Figure S6: RMSF of glycan trees of AGP (black with fucoses and red without fucoses).

5 Conclusões

A partir dos objetivos traçados, o presente trabalho permitiu:

- Reprodução, em mais de 95%, do comportamento conformacional da parte sacarídica de glicoproteínas e dissacarídeos em solução aquosa, em comparação a dados experimentais de RMN, validando o protocolo empregado para descrição de carboidratos por MM;
- Utilização do refinamento de mínimos de energia obtidos por mapas de contorno em solução aquosa por DM como uma estratégia capaz de agregar qualidade à descrição conformacional de carboidratos, mais próxima do que deve ser esperado em soluções biológicas e utilizando o campo de força GROMOS96 43a1;
- Proposição do emprego de estados conformacionais de dissacarídeos como uma metodologia onde os dissacarídeos funcionam como blocos para a construção de glicanas complexas ligadas a glicoproteínas;
- Utilização da metodologia acima para construção de glicoproteínas sem estrutura tridimensional glicídica identificada;
- Entendimento do papel de glicosilação na estrutura da AGP e as implicações que a glicosilação tem sobre o sitio de ligação da proteína.
- A partir da glicosilação da AGP foi possível a caracterização do complexo AGP - P-selectina, propondo uma forma de ligação entre as duas moléculas e algumas bases moleculares dessa ligação a duas árvores glicídicas diferentes.

De forma geral, os resultados obtidos confirmam o potencial de ferramentas de MM no estudo de sistemas biológicos, especialmente na reprodução e predição de propriedades conformacionais de oligossacarídeos, correlacionando-se com dados experimentais prévios e possibilitando a construção de estruturas glicosiladas de proteínas a partir da estrutura protéica e da sequência de glicosilação.

6 Referências Bibliográficas

Referências citadas no texto e nos artigos publicados

ABU-QARN, M.; EICHLER, J.; SHARON, N.: Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.*, **2008**, *18*, 544-550.

ALENCAR, N. M. N.; OLIVEIRA, R. S. B.; FIGUEIREDO, J. G.; CAVALCANTE, I. J. M.; MATOS, M. P. V.; CUNHA, F. Q.; NUNES, J. V. S.; BOMFIM, L. R.; RAMOS, M. V.: An anti-inflammatory lectin from Luetzelburgia auriculata seeds inhibits adhesion and rolling of leukocytes and modulates histamine and PGE2 action in acute inflammation models, *Inflammation Research*, **2009**, *59*, 245-254.

ALI, M. M.; AICH, U.; VARGHESE, B.; PÉREZ, S.; IMBERTY, A.; LOGANATHAN, D.: Conformational Preferences of the Aglycon Moiety in Models and Analogs of GlcNAc-Asn Linkage: Crystal Structures and ab Initio Quantum Chemical Calculations of N-(β-D-Glycopyranosyl)haloacetamides. *J. Am. Chem. Soc.*, **2008**, *130*, 8317-8325.

ALLINGER, N. L.: Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.*, **1977**, 99, 8127-8134.

ALLINGER, N. L.; YUH, Y. H.; LII, J. -H.: Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, **1989**, *111*, 8551-8567.

ALLINGER, N. L.; RAHMAN, M.; LII, J. -H.: A molecular mechanics force field (MM3) for alcohols and ethers. *J. Am. Chem. Soc.*, **1990**, *112*, 8293-8307.

ALMOND, A.; DEANGELIS, P. L.; BLUNDELL, C. D.: Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix. *J. Mol. Biol.*, **2006**, *358*, 1256-1269.

ANDREC, M.; SNYDER, D. A.; ZHOU, Z.; YOUNG, J.; MONTELIONE, G. T.; LEVY, R. M.: A large data set comparison of protein structures determined by crystallography and NMR: Statistical test for structural differences and the effect of crystal packing. *Proteins*, **2007**, *69*, 449-465.

ARULANANDAM, A. R. N.; WITHKA, J. M.; WYSS, D. F.; WAGNER, G.; KISTER, A.; PALLAI, P.; RECNY, M. A.; REINHERZ, E. L.: The CD58 (LFA-3) binding site is a localized and highly charged surface area on the AGFCC'C" face of the

human CD2 adhesion domain. *Proc. Natl. Acad. Sci. U.S.A.*, **1993**, *90*, 11613-11617.

BANNER, D. W.; D'ARCY, A.; CHÈNE, C.; WINKLER, F. K.; GUHA, A.; KONIGSBERG, W. H.; NEMERSON, Y.; KIRCHHOFER, D.: The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, **1996**, *380*, 41-46.

BARTHE, P.; PUJADE-RENAUD, V.; BRETON, F.; GARGANI, D.; THAI, R.; ROUMESTAND, C.; DE LAMOTTE, F.: Structural Analysis of Cassiicolin, a Host-selective Protein Toxin from *Corynespora cassiicola*. *J. Mol. Biol.*, **2007**, *367*, 89-101.

BECKER, C. F.; GUIMARÃES, J. A.; VERLI, H.: Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution. *Carbohydr. Res.*, **2005**, *340*, 1499-1507.

BECKER, C. F.; GUIMARÃES, J. A.; MOURÃO, P. A. S.; VERLI, H.: Conformation of sulfated galactan and sulfated fucan in aqueous solutions: Implications to their anticoagulant activities. *J. Mol. Graph. Model.*, **2007**, *26*, 391-399.

BERENDSEN, H. J. C.; POSTMA, J. P. M.; DINOLA, A.; HAAK, J. R.: Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, **1984,** *81,* 3684-3690.

BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P.: The missing term in effective pair potentials. *J. Phys. Chem.*, **1987**, *91*, 6269-6271.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P.E.: The protein data bank. Nucleic Acids Res. **2000**, 28, 235-242.

BERMAN, H. M.; HENRIC, K.; NAKAMURA, H.; MARKLEY, J.: Building meaningful models of glycoproteins. *Nat. Struct. Mol. Biol.*, **2007**, *14*, 354-355.

BLUHM, T. L.; DESLANDES, Y.; MARCHESSAULT, R. H.; PEREZ, S.; RINAUDO, M.: Solid-state and solution conformation of scleroglucan. *Carbohydr. Res.*, **1982**, *100*, 117-130.

BOCK, K.: The preferred conformation of oligosaccharides in solution inferred from high resolution NMR data and hard sphere exo-anomeric calculation. *Pure & Appi. Chem.,* **1983**, 55, 605-622,

BOHNE-LANG, A.; VON DER LIETH, C. W.: Glycosylation of proteins: a computer based method for the rapid exploration of conformational space of N-glycans. *Pac. Symp. Biocomput.*, **2002**, *7*, 285-296.

BOSQUES, C. J.; TSCHAMPEL, S. M.; WOODS, R. J.; IMPERIALI, B.: Effects of Glycosylation on Peptide Conformation: A Synergistic Experimental and Computational Study. *J. Am. Chem. Soc.*, **2004**, *126*, 8421-8425.

BRANDEN C., TOOZE, J.: Introduction to protein structure. 2.ed. New York: Garland, **1999**. p. 373-392.

BRANT, D. A.: Shapes and motions of polysaccharide chains. *Pure Appl. Chem.*, **1997**, *69*, 1885-1892.

CASE, D. A.; CHEATHAM, T. E. 3^{RD}; DARDEN, T.; GOHLKE, H.; LUO, R.; MERZ, K. M. JR.; ONUFRIEV, A.; SIMMERLING, C.; WANG, B.; WOODS, R. J.: The Amber Biomolecular Simulation Programs. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.

CASTRO, M. O.; POMIN, V. H.; SANTOS, L. L.; VILELA-SILVA, A.-C. E. S.; HIROHASHI, N.; POL-FACHIN, L.; VERLI, H.; MOURÃO, P. A. S.: A Unique 2-Sulfated â-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.

CARUGO, O.; ARGOS, P.: Protein-protein crystal-packing contacts. Protein Sci. **1997**, 6, 2261-2263.

CECILIANI, F.; POCACQUA, V.: The Acute Phase Protein alpha1-Acid Glycoprotein: A Model for Altered Glycosylation During Diseases, *Current Protein and Peptide Science*, **2007**, *8*, 91-108.

CHAVEZ, M. I.; ANDREU, C.; VIDAL, P.; ABOITIZ, N.; FREIRE, N.; GROVES, P.; ASENSIO, J. L; ASENSIO, G.; MURAKI, M.; CANADA, F. J.; JIMENEZ-BARBERO, J.: Carbohydrate-aromatic interactions for the molecular recognition of oligosaccharides by proteins: NMR studies of the structure and binding affinity of AcAMP2-like peptides with non-natural naphthyl and fluoroaromatic residues. *Chemistry*, 2005, 11, 7060-7074.

CIPOLLO, J. F.; AWAD, A. M.; COSTELLO, C. E.; HIRSCHBERG, C. B.: *N*-Glycans of *Caenorhabditis elegans* Are Specific to Developmental Stages. *J. Biol. Chem.*, **2005**, *280*, 26063-26072.

CLARK, M.; CRAMER, R. D.; VAN OPDENBOSCH, N.: Validation of the general-purpose Tripos 5.2 force-field. *J. Comput. Chem.*, **1989**, *10*, 982-1012.

COPIÉ, V.; TOMITA, Y.; AKIYAMA, S. K.; AOTA, S.; YAMADA, K. M.; VENABLE, R. M.; PASTOR, R. W.; KRUEGER, S.; TORCHIA, D. A.: Solution Structure and Dynamics of Linked Cell Attachment Modules of Mouse Fibronectin Containing the RGD and Synergy Regions: Comparison with the Human Fibronectin Crystal Structure. *J. Mol. Biol.*, **1998**, 277, 663-682.

CORZANA, F.; MOTAWIA, M. S.; DU PENHOAT, C. H.; PÉREZ, S.; TSCHAMPEL, S. M.; WOODS, R. J.; ENGELSEN, S. B.: A Hydration Study of (1→4) and (1→6) Linked α-Glucans by Comparative 10 ns Molecular Dynamics Simulations and 500-MHz NMR. *J. Comput. Chem.*, **2003**, *25*, 573-586.

CORZANA, F.; BUSTO, J. H.; JIMÉNEZ-OSÉS, G.; ASENSIO, J. L.; JIMÉNEZ-BARBERO, J.; PEREGRINA, J. M.; AVENOZA, A.: New Insights into α-GalNAc–Ser Motif: Influence of Hydrogen Bonding versus Solvent Interactions on the Preferred Conformation. *J. Am. Chem. Soc.*, **2006**, *128*, 14640-14648.

COULDREY, C.; GREEN, J. E.: Metastases: the glycan connection. *Breast Cancer Res.*, **2000**, *2*, 321-323.

CRISPIN, M.; STUART, D. I.; JONES, E. Y.: Building meaningful models of glycoproteins. *Nat. Struct. Mol. Biol.*, **2007**, *14*, 354.

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

CUMMING, D. A.; CARVER, J. P.: Reevaluation of rotamer population for 1,6 linkages: reconcilation with potential energy calculations. *Biochemistry*, **1987**, *26*, 6676-6683.

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.

DE BEER, T.; VLIEGENTHART, J. F.; LÖFFLER, A.; HOFSTEENGE, J.: The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase U_S Is α-Mannopyranose. *Biochemistry*, **1995**, *34*, 11785-11789.

DE GROOT, B. L.; GRUBMÜLLER, H.: Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science*, **2001**, *294*, 2353-2357.

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

DELL, A; MORRIS, H. R.: Glycoprotein Structure Determination by Mass Spectrometry. *Science*, **2001**, 291, 2351-2356.

DELANO, W. L.: The PyMOL Molecular Graphics System, DeLano Scientific, Sao Carlos, CA, USA, 2002. http://www.pymol.org

DIMITROPOULOS, N.; PAPAKYRIAKOU, A.; DALKAS, G. A.; CHASAPIS, C. T.; POULAS, K.; SPYROULIAS G. A.: A computational investigation on the role of glycosylation in the binding of alpha1 nicotinic acetylcholine receptor with two alpha-neurotoxins. *Proteins: Structure, Function, and Bioinformatics***, 2011,** 79, 142–152.

DOUCEY, M. A.; HESS, D.; CACAN, R.; HOFSTEENGE, J.: Protein C-Mannosylation Is Enzyme-catalyzed and Uses Dolichyl-Phosphate-Mannose as Precursor. *Mol. Biol. Cell*, **1998**, *9*, 291-300.

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

ERBEL, P. J. A.; KARIMI-NEJAD, Y.; DE BEER, T.; BOELENS, R.; KAMERLING, J. P.; VLIEGENTHART, J. F. G.: Solution structure of the α-subunit of human chorionic gonadotropin. *Eur. J. Biochem.*, **1999**, *260*, 490-498.

ERBEL, P. J. A.; KARIMI-NEJAD, Y.; VAN KUIK, J. A.; BOELENS, R.; KAMERLING, J. P.; VLIEGENTHART, J. F. G.: Effects of the N-Linked Glycans on the 3D Structure of the Free α-Subunit of Human Chorionic Gonadotropin. *Biochemistry*, **2000**, *39*, 6012-6021.

EYAL, E.; GERZON, S.; POTAPOV, V.; EDELMAN, M.; SOBOLEV, V.: The Limit of Accuracy of Protein Modeling: Influence of Crystal Packing on Protein Structure. *J. Mol. Biol.*, **2005**, *351*, 431-442.

FITOS, I.; VISY, J.; ZSILA, F.; MADY, G.; SIMONYI, M.: Selective binding of imatinib to the genetic variants of human α1-acid glycoprotein, *Biochim. Biophys. Acta* **2006**, *1760*, 1704-1712.

FLETCHER, C. M.; HARRISON, R. A.; LACHMANN, P. J.; NEUHAUS, D.: Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure*, **1994**, *2*, 185-199.

FOURNIER, T.; MEDJOUBI-N, N.; PORQUET, D.: Alpha-1-acid glycoprotein, *Biochimica et Biophysica Acta (BBA) - General Subjects,* **2000**, *1482*, 157-171.

FOXALL, C.; WATSON, S. R.; DOWBENKO, D.; FENNIE, C.; LASKY, L. A.; KISO, M.; HASEGAWA, A.; ASA, D.; BRANDLEY, B. K.: The three of the selectin receptor family recognize a common carbohydrate epitope, the sialyl Lewis X oligosaccharide. *J. Cell Biol.*, **1992**, *117*, 895–902.

FUKUDA, M.: Leukosialin, a major O-glycan-containing sialoglycoprotein defining leukocyte differentiation and malignancy. *Glycobiology*, **1991**, *1*, 347-356.

FURSE, K. E.; PRATT, D. A.; PORTER, N. A.; LYBRAND, T. P.: Molecular Dynamics Simulations of Arachidonic Acid Complexes with COX-1 and COX-2: Insights into Equilibrium Behavior. *Biochemistry*, **2006**, *45*, 3189-3205.

FUSE, E.; TANII, H.; TAKAI, K.; ASANOME, K.; KURATA, N.; KOBAYASHI, H.; SUGIYAMA, Y.: Altered pharmacokinetics of a novel anticancer drug, UCN-01, caused by specific high affinity binding to α1-acid glycoprotein in humans. *Cancer Res.*, **1999**, 59, 1054–1060.

GRAAF, T. W. D.; STELT, M. E. V. D.; ANBERGEN, M. G.; DIJK, W. V.: Inflammation-induced Expression of Sialyl Lewis X-containlng Glycan Structures on alpha-1-Acid Glycoprotein (Orosomucoid) in Human Sera, *J. Exp. Med.,* **1993**, *177*, 657-666.

GUNNARSSON, P.; LEVANDER, L.; PÅHLSSON, P.; GRENEGÅRD, M.: The acute-phase protein alpha1-acid glycoprotein (AGP) induces rises in cytosolic Ca2+ in neutrophil granulocytes *via* sialic acid binding immunoglobulinlike lectins (Siglecs). *The FASEB journal*, **2007**, 21, 4059-4069.

GUPTA, R.; BIRCH, H.; RAPACKI, K.; BRUNAK, S.; HANSEN, J. E.: O-GLYCBASE version 4.0: a revised databse of O-glycosylated proteins. *Nucleic Acids Res.*, **1999**, *27*, 370-372.

HA, S. N.; GIAMMONA, A.; FIELD, M.; BRADY, J. W.: A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr. Res.*, **1988**, *180*, 207-221.

HANSEN, A. P.; PETROS, A. M.; MEADOWS, R. P.; NETTESHEIM, D. G.; MAZAR, A. P.; OLEJNICZAK, E. T.; XU, R. X.; PEDERSON, T. M.; HENKIN, J.; FESIK, S. W.: Solution structure of the amino-terminal fragment of urokinase-type plasminogen activator. *Biochemistry*, **1994**, *33*, 4847-4864.

HASHIMOTO, Y.; TOMA, K.; NISHIKIDO, J.; YAMAMOTO, K.; HANEDA, K.; INAZU, T.; VALENTINE, K. G.; OPELLA, S. J.: Effects of Glycosylation on the Structure and Dynamics of Eel Calcitonin in Micelles and Lipid Bilayers Determined by Nuclear Magnetic Resonance Spectroscopy. *Biochemistry*, **1999**, *38*, 8377-8384.

HASHIMOTO, K.; GOTO, S.; KAWANO, S.; AOKI-KINOSHITA, K. F., UEDA, N.; HAMAJIMA, M.; KAWASAKI, T.; KANEHISA, M.: KEGG as a glycome informatics resource. *Glycobiology*, **2006**, *16*, 63R-70R.

HAYNES, P. A.: Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology*, **1998**, 8, 1-5.

HELENIUS, A.; AEBI, M.: Intracellular Functions of N-Linked Glycans. *Science*, **2001**, *291*, 2364-2369.

HEMMINGSEN, L.; MADSEN, D. E.; ESBENSEN, A. L.; OLSEN, L.; ENGELSEN, S. B.: Evaluation of carbohydrate molecular mechanical force fields by quantum mechanical calculations. *Carbohydr. Res.*, **2004**, *339*, 937-948.

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.

HOFSTEENGE, J.; MÜLLER, D. R.; DE BEER, T.; LÖFFLER, A.; RICHTER, W. J.; VLIEGENTHART, J. F.: New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase U_S. *Biochemistry*, **1994**, *33*, 13524-13530.

HOSOGUCHI, K.; SHIMIZU, K.; FUJITANI, N.; NISHIMURA, S.: Chemical Synthesis, Folding, and Structural Insights into O-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor *J.Am.Chem.Soc.*,**2010**, 132, 14857-14865.

HUANG, X.; BARCHI, J. J. JR.; LUNG, F. D.; ROLLER, P. P.; NARA, P. L.; MUSCHIK, J.; GARRITY, R. R.: Glycosylation affects both the three-dimensional structure and antibody binding properties of the HIV-1IIIB GP120 peptide RP135. *Biochemistry*, **1997**, *36*, 10846-10856.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K.: VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, **1996,** *14,* 33-38. http://www.ks.uiuc.edu/Research/vmd/.

HWANG, M. J.; NI, X.; WALDMAN, M.; EWIG, C. S.; HAGLER, A. T.: Derivation of class II force fields. VI. Carbohydrate compounds and anomeric effects. *Biopolymers*, **1998**, *45*, 435-468.

IMBERTY, A.; PÉREZ, S.: Stereochemistry of the N-glycosylation sites in glycoproteins. *Protein Eng.*, **1995**, *8*, 699-709.

IMBERTY, A.; PÉREZ, S.: Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem. Rev.*, **2000**, 100, 4567–4588.

IMPERIALI, B.; HENDRICKSON, T. L.: Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.*, **1995**, *3*, 1565-1578.

IMREA, T.; KREMMERB, T.; HÉBERGERB, K.; MOLNÁR-SZÖLLŐSIC, E.; LUDÁNYIA, K.; PÓCSFALVID, G.; MALORNID, A.; DRAHOSA, L.; VÉKEY, K.: Mass spectrometric and linear discriminant analysis of N-glycans of human serum alpha-1-acid glycoprotein in cancer patients and healthy individuals. *Journal of proteomics*, **2008**, 71, 186-197.

IUPAC-IUB Commission on Biochemical Nomenclature: Conformational nomenclature for five- and six-membered ring forms of monosaccharides and their derivatives (Recommendations 1980). *Eur. J. Biochem.*, **1980**, *111*, 295-298.

IUPAC-IUB Commission on Biochemical Nomenclature: Symbols for specifying the conformation of polysaccharide chains. *Pure Appl. Chem.*, **1983**, *55*, 1269-1272.

IUPAC-IUB Commission on Biochemical Nomenclature: Nomenclature for carbohydrates. *Pure Appl. Chem.*, **1996**, *68*, 1919-2008.

JAEKEN, J.; CARCHON, H.: The carbohydrate-deficient glycoprotein syndromes: An overview. *J. Inher. Metab. Dis.*, **1993**, *16*, 813-820.

JAIN, S.; SUNDARALINGAM, M.: Effect of Crystal Packing Environment on Conformation of the DNA Duplex. *J. Biol. Chem.*, **1989**, *264*, 12780-12784.

JORGENSEN, W. L.; TIRADO-RIVES, J.: The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, **1988**, *110*, 1657-1666.

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

KANSAS, G.S.: Selectins and their ligands: current concepts and controversies. *Blood*, **1996**, *88*, 3259–3287.

KAO, Y. H.; LEE, G. F.; WANG, Y.; STAROVASNIK, M. A.; KELLEY, R. F.; SPELLMAN, M. W.; LERNER, L.: The Effect of O-Fucosylation on the First EGF-like Domain from Human Blood Coagulation Factor VII. *Biochemistry*, **1999**, *38*, 7097-7110.

KARPLUS, M.; PETSKO, G. A.: Molecular dynamics simulations in biology. *Nature*, **1990**, *347*, 631-639.

KIRSCHNER, K. N.; YONGYE, A. B.; TSCHAMPEL, S. M.; GONZÁLEZ-OUTEIRIÑO, J.; DANIELS, C. R.; FOLEY, B. L.; WOODS, R. J.: GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.*, **2008**, *29*, 622-655.

KITSON, D. H.; HAGLER, A. T.: Theoretical-studies of the structure and molecular-dynamics of a peptide crystal. *Biochemistry*, **1988**, *27*, 5246-5257.

KODICEK, M.; INFANZON, A.; KARPENKO, V.: Heat denaturation of human orosomucoid in water/methanol mixtures. *Biochim. Biophys. Acta*, **1995**, 1246, 10–16.

KRÄUTLER, V.; MÜLLER, M.; HÜNENBERGER, P. H.: Conformation, dynamics, solvation and relative stabilities of selected beta-hexopyranoses in water: a molecular dynamics study with the GROMOS 45A4 force field. *Carbohydr. Res.*, **2007**, *342*, 2097-2124.

KREMER, J. M.; WILTING, J.; JANSSEN, L. H.: Drug binding to human alpha-1-acid glycoprotein in health and disease. *Pharmacol. Rev.*, **1988**, 40, 1–47.

KRIEG, J.; HARTMANN, S.; VICENTINI, A.; GLÄSNER, W.; HESS, D.; HOFSTEENGE, J.: Recognition Signal for C-Mannosylation of Trp-7 in RNase 2 Consists of Sequence Trp-x-x-Trp. *Mol. Biol. Cell*, **1998**, *9*, 301-309.

KUTTEL, M. M.; NAIDOO, K. J.: Free Energy Surfaces for the α(1-4)-Glycosidic Linkage: Implications for Polysaccharide Solution Structure and Dynamics. *J. Phys. Chem. B*, **2005**, *109*, 7468-7474.

LASKY, L. A.: Selectins: interpreters of cell-specific carbohydrate information during inflammation, *Science,* **1992**, *258*, 964-969.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M.: PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **1993,** *26,* 283-291.

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

LEMIEUX, R. U.; BOCK, K.: The conformational analysis of oligosaccharides by ^{1}H-NMR and HSEA calculation. *Arch Biochem Biophys.*, **1983**, 221,125-134.

LEVANDER, L.; GUNNARSSON, P.; GRENEGÅRD, M.; RYDÉN, I.; PÅHLSSON, P.: Effects of α1-acid Glycoprotein Fucosylation on its Ca2+Mobilizing Capacity in Neutrophils, *Scandinavian Journal of Immunology,* **2009**, *69*, 412-420.

LI, S.; LIU, B.; ZENG, R.; CAI, Y.; LI, Y.: Predicting O-glycosylation sites in mammalian proteins by using SVMs. *Comput. Biol. Chem.*, **2006**, *30*, 203-208.

LINDEN, E. C. M. B.-V. D.; OMMEN, E. C. R. V.; DIJK, W. V.: Glycosylation of alpha-acid glycoprotein in septic shock: changes in degree of branching and in expression of sialyl Lewis" groups, *Glycoconjugate Journal,* **1996**, *13*, 27-31.

LINS, R. D.; HÜNENBERGER, P. H.: A New GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.*, **2005**, *26*, 1400-1412.

LIPKIND, G. M.; VEROVSKY, V. E.; KOCHETKOV, N. K.: Conformational states of cellobiose and maltose in solution: A comparison of calculated and experimental data. *Carbohydr. Res.*, **1984**, *133*, 1-13.

LIPKIND, G. M.; SHASHKOV, A. S.; KOCHETKOV, N. K.: Nuclear overhauser effect and conformational states of cellobiose in aqueous solution. *Carbohydr. Res.*, **1985**, *141*, 191-197.

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

LÜTTEKE, T.; FRANK, M.; VON DER LIETH, C. -W.: Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydr. Res.*, **2004**, *339*, 1015-1020.

LÜTTEKE, T., FRANK, M.; VON DER LIETH, C. -W.: Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res.*, **2005**, *33*, D242-D246.

LÜTTEKE, T.; BOHNE-LANG, A.; LOSS, A.; GOETZ, T.; FRANK, M.; VON DER LIETH, C. -W.: GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology*, **2006**, *16*, 71R-81R.

LÜTTEKE, T.: Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr. D Biol. Crystallogr.*, **2009**, *65*, 156-168.

MACKERELL, 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. 3^{RD}; 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*, **1998**, *102*, 3586-3616.

MACKERELL, A. D. JR.: Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, **2004**, *25*, 1584-1604.

MANDAL, T. K.; MUKHOPADHYAY, C.: Effect of Glycosylation on Structure and Dynamics of MHC Class I Glycoprotein: A Molecular Dynamics Study. *Biopolymers*, **2001**, *59*, 11-23.

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

MARSHALL, R. D.: Glycoproteins. *Annu. Rev. Biochem.*, **1972**, *41*, 673-702.

MARTÍN-PASTOR, M.; ESPINOSA, J. F.; ASENSIO, J. L.; JIMÉNEZ-BARBERO, J.: A comparison of the geometry and of the energy results obtained by application of different molecular mechanics force fields to methyl *α*-lactoside and the C-analogue of lactose. *Carbohydr. Res.*, **1996**, *298*, 15-49.

MEDVEDOVÁ, L.; FARKAS, R.: Hormonal control of protein glycosylation: role of steroids and related lipophilic ligands. *Endocr. Regul.*, **2004**, *39*, 65-79.

MEHTA, D. P.; ICHIKAWA, M.; SALIMATH, P. V.; ETCHITSON, J. R.; HAAK, R.; MANZI, A.; FREEZE, H. H.: A lysosomal cysteine proteinase from *Dictyostelium discoideum* contains N-acetylglucosamine 1-phosphate bound to serine but not mannose-6-phosphate on N-linked oligosaccharides. *J. Biol. Chem.*, **1996**, 271, 10897-10903.

METZLER, W. J.; BAJORATH, J.; FENDERSON, W.; SHAW, S. Y.; CONSTANTINE, K. L.; NAEMURA, J.; LEYTZE, G.; PEACH, R. J.; LAVOIE, T. B.; MUELLER, L.; LINSLEY, P. S.: Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. *Nat. Struct. Biol.*, **1997**, *4*, 527-531.

MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M.: Dynamics of folded proteins. *Nature*, **1977**, *267*, 585-590.

MORRIS, G. M.; GOODSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R. K.; OLSON, A. J.: Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function, *Journal of Computational Chemistry,* **1998**, *19*, 1639-1662.

MUKHOPADHYAY, C.: Molecular Dynamics Simulation of Glycoprotein-Glycans of Immunoglobulin G and Immunoglobulin M. *Biopolymers*, **1998**, *45*, 177-190.

MULLOY, B.; FORSTER, M. J.; JONES, C.; DAVIES, D. B.: N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.*, **1993**, *293*, 849-858.

MUTSAERS, J. H. G. M.; VAN HALBEEK, H.; KAMERLING, J. P.; VLIEGENTHART, J. F. G.: Determination of the structure of the carbohydrate chains of prostaglandin endoperoxide synthase from sheep. *Eur. J. Biochem.*, **1985**, *147*, 569-574.

NAGY, K.; VÉKEY, K.; IMRE, T.; LUDÁNYI, K.; BARROW, M. P.; DERRICK, P. J.: Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Human α-1-Acid Glycoprotein, *Analytical Chemistry,* **2004**, *76*, 4998-5005.

NAIDOO, K. J.; DENYSYK, D.; BRADY, J. W.: Molecular dynamics simulations of the N-linked oligosaccharide of the lectin from *Erythrina corallodendron*. *Protein Eng.*, **1997**, *10*, 1249-1261.

NAKANO, M.; KAKEHI, K.; TSAI, M.-H.; LEE, Y. C.: Detailed structural features of glycan chains derived from a1-acid glycoproteins of several different animals: the presence of hypersialylated, O-acetylated sialic acids but not disialyl residues, *Glycobiology,* **2004**, *14*, 431-441.

NAVARRO, A. S. ; DIAS, S. M. G. ; MELLO, L. V. ; GIOTTO, M. T. S.; BLONSKI, C. ; GARRAT, R. C. ; RIGDEN, D. J.: Structural flexibility in Trypanosoma brucei enolase revealed by X-ray crystallography and molecular dynamics. *The FEBS Journal*, **2007,** 274, p. 5077-5089.

NEMETH, J. F.; HOCHENSANG, G. P.; MARNETT, L. J.; CAPRIOLI, R. M.: Characterization of the glycosylation sites in cyclooxygenase-2 using mass spectrometry. *Biochemistry*, **2001**, *40*, 3109-3116.

NGUYEN, D. H.; COLVIN, M. E.; YEH, Y.; FEENEY, R. E.; FINK, W. H.: The Dynamics, Structure, and Conformational Free Energy of Proline-Containing Antifreeze Glycoprotein. *Biophys. J.*, **2002**, *82*, 2892-2905.

OLEWICZ-GAWLIK, A.; KORCZOWSKA-LACKA, I.; LACKI, J. K.; KLAMA, K.; HRYCAJ, P.: Fucosylation of serum α1-acid glycoprotein in rheumatoid arthritis patients treated with infliximab, *Clinical Rheumatology, **2007**, 26*, 1679-1684.

OTTO, J. C.; DEWITT, D. L.; SMITH, W. L.: N-glycosylation of prostaglandin endoperoxide synthases-1 and -2 and their orientations in the endoplasmic reticulum. *J. Biol. Chem.*, **1993**, *268*, 18234-18242.

PARIKH, H. H.; MCELWAIN, K.; BALASUBRAMANIAN, V.; LEUNG, W.; WONG, D.; MORRIS, M. E.; RAMANATHAN, M.: A rapid spectrofluorimetric technique for determining drug–serum protein binding suitable for high-throughput screening, *Pharm. Res.*, **2000**, *17*, 632-637.

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

PERERA, L.; DARDEN, T. A.; PEDERSEN, L. G.: Predicted Solution Structure of Zymogen Human Coagulation FVII. *J. Comput. Chem.*, **2002**, *23*, 35-47.

PÉREZ, S.; IMBERTY, A.; ENGELSEN, S. B.; GRUZA, J.; MAZEAU, K.; JIMENEZ-BARBERO, J.; POVEDA, A.; ESPINOSA, J.-F.; VAN EYCK, B. P.; JOHNSON, G.; FRENCH, A. D.; KOUWIJZER, M. L. C. E.; GROOTENUIS, P. D. J.; BERNARDI, A.; RAIMONDI, L.; SENDEROWITZ, H.; DURIER, V.; VERGOTEN, G.; RASMUSSEN, K.: A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr. Res*., **1998**, *314*, 141-155.

PÉREZ, S.; MULLOY, B.: Prospects for glycoinformatics. *Curr. Opin. Struc. Biol.*, **2005**, *15*, 517-524.

PETERSON, A.; SEED, B.: Monoclonal antibody and ligand binding sites of the T cell erythrocyte receptor (CD2). *Nature*, **1987**, *329*, 842-846.

PETRESCU, A. J.; PETRESCU, S. M.; DWEK, R. A.; WORMALD, M. R.: A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology*, **1999**, *9*, 343-352.

PETRESCU, A. J.; MILAC, A. L.; PETRESCU, S. M.; DWEK, R. A.; WORMALD, M. R.: Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology*, **2004**, *14*, 103-114.

PETRESCU, A. J.; WORMALD, M. R.; DWEK, R. A.: Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr. Opin. Struct. Biol.*, **2006**, *16*, 600-607.

PHILIPPOPOULOS, M.; LIM, C.: Exploring the Dynamic Information Content of a Protein NMR Structure: Comparison of a Molecular Dynamics Simulation With the NMR and X-Ray Structures of *Escherichia coli* Ribonuclease HI. Proteins. **1999**, 36, 87-110.

PICKFORD, A. R.; SMITH, S. P.; STAUNTON, D.; BOYD, J.; CAMPBELL, I. D.: The hairpin structure of the ^{6}F1^{1}F2^{2}F2 fragment from human fibronectin enhances gelatin binding. *EMBO J.*, **2001**, *20*, 1519-1529.

POL-FACHIN, L.; VERLI, H.: Depiction of the forces participating in the 2-*O*-sulfo-α-L-iduronic acid conformational preference in heparin sequences in aqueous solutions. *Carbohydr. Res.*, **2008**, *343*, 1435-1445.

POL-FACHIN, L.; FERNANDES, C. L.; VERLI, H.: GROMOS96 43a1 performance on the characterization of glycoproteins conformational ensemble through molecular dynamics simulations. *Carbohydr. Res.* **2009**, *344*, 491-500.

POL-FACHIN, L; VERLI, H.: Assessment of glycoproteins dynamics from computer simulations. *Mini-Reviews in Organic Chemistry*, **2011**, no prelo.

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

POPE, L.: Modeling carbohydrate conformations from NMR data: maximum entropy rotameric distribution about the C5-C6 bond in gentiobiose. *J. Am. Chem. Soc.*, **1993**, *115*, 8421-8426.

RECNY, M. A.; LUTHER, M. A.; KNOPPERS, M. H.; NEIDHARDT, E. A.; KLANDEKAR, S. S.; CONCINO, M. F.; SCHIMKE, P. A.; FRANCIS, M. A.; MOEBIUSLL, U.; REINHOLD, B. B.; REINHOLD, V. N.; REINHERZ, E. L.: N-Glycosylation Is Required for Human CD2 Immunoadhesion Functions. *J. Biol. Chem.*, **1992**, *267*, 22428-22434.

RESS, D. A.; SCOTT, W. E.: Polysaccharide conformation. Part VI. *J. Chem. Soc. B*, **1971**, 469-479.

RIGBY, A. C.; LUCAS-MEUNIER, E.; KALUME, D. E.; CZERWIEC, E.; HAMBE, B.; DAHLQVIST, I.; FOSSIER, P.; BAUX, G.; ROEPSTORFF, P.; BALEJA, J. D.; FURIE, B. C.; FURIE, B.; STENFLO, J.: A conotoxin from Conus textile with unusual posttranslational modifications reduces presynaptic Ca2+ influx. *Proc. Natl. Acad. Sci. U. S. A.*, **1999**, *96*, 5758-5763.

ROJO-DOMIGUEZ, A.; HERNANDEZ-ARANA, A.: Three-dimensional modeling of the protein moiety of human alpha1-acid glycoprotein, a lipocalin-family member. *Protein Sequences Data Anal*., **1993**, 5, 349-355.

ROLLINS, T. E.; SMITH, W. L.: Subcellular localization of prostaglandin-forming cyclooxygenase in Swiss mouse 3T3 fibroblasts by electron microscopic immunocytochemistry. *J. Biol. Chem.*, **1980**, *255*, 4872-4875

RUBINSTEIN, A.; KINARSKY, L.; SHERMAN, S.: Molecular Dynamics Simulations of the O-glycosylated 21-residue MUC1 Peptides. *Int. J. Mol. Sci.*, **2004**, *5*, 119-128.

RUTHERFORD, T. J.; PARTRIDGE, J.; WELLER, C. T.; HOMANS, S. W.: Characterization of the extent of internal motions in oligosaccharides. *Biochemistry*, **1993**, 32, 12715-12724.

SÁEZ, F. J.; MADRID, J. F.; ALONSO, E.; HERNÁNDEZ, F.: Glycan composition of follicle (Sertoli) cells of the amphibian *Pleurodeles waltl*. A lectin histochemical study. *J. Anat.*, **2001**, *198*, 673-681.

SANBONMATSU, K. Y.; JOSEPH, S.; TUNG, C. S.: Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. U. S. A*., **2005**, *102*, 15854-15859.

SATHYANARAYANA, B. K.; RAO, V. S. R.: Conformational studies of -glucans. *Biopolymers*, **1972**, *11*, 1379-1394.

SATHYANARAYANA, B. K.; STEVENS, E. S.: Theoretical study of the conformations of pustulan [(1----6)-beta-D-glucan]. *J. Biomol. Struct. Dyn.*, **1983**, *1*, 947-959.

SCHAFTENAAR, G.: MOLDEN. CAOS/CAMM Center, University of Nijmegen, Toernooiveld 1, 6525 ED NIJMEGEN, The Netherlands, 1997.

SCHAFTENAAR, G.; NOORDIK, J. H.: MOLDEN: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 123-134.

SCHLICK, T.: Molecular Modeling and Simulation: an Interdisciplinary Guide, 1.ed. New York: Springer, **2006**.

SCHMID, K.: Preparation and properties of an acid glycoprotein prepared from human plasma, *Journal of the american chemical society* , **1950**, *72*, 2816-2816.

SCHMID, K.: Preparation and Properties of Serum and Plasma Proteins. XXM. Separation from Human Plasma of Polysaccharides, Peptides and Proteins of Low Molecular Weight. Crystallization of an Acid Glycoprotein, *Journal of the american chemical society, **1953**, 75*, 60-68.

SCHMID, K.; BURGI, W.; COLLINS, J. H.; NANNO, S.: The Disulfide Bonds of alpha1-Acid Glycoprotein. *Biochemistry*, **1974**, 13, 2694-2697.

SCHMIDT, M. A.; RILEY, L. W.; BENZ, I.: Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.*, **2003**, *11*, 554-561.

SCHONFELD, D.; RAVELLI, R.; MUELLER, U.; SKERRA, A.: The 1.8-Å Crystal Structure of α1-Acid Glycoprotein (Orosomucoid) Solved by UV RIP Reveals the Broad Drug-Binding Activity of This Human Plasma Lipocalin, *Journal of Molecular Biology,* **2008**, *384*, 393-405.

SCHÜTTELKOPF, A. W.; VAN AALTEN, D. M. F.: PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.*, **2004**, D60, 1355-1363.

SEI, K.; NAKANO, M.; KINOSHITA, M.; MASUKO, T.; KAKEHI, K.: Collection of a1-acid glycoprotein molecular species by capillary electrophoresis and the analysis of their molecular masses and carbohydrate chains Basic studies on the analysis of glycoprotein glycoforms, *Journal of Chromatography A,* **2002**, *958*, 273-271.

SHEFTER, E.; TRUEBLOOD, K. N.: The crystal and molecular structure of D(+)-barium uridine-5'-phosphate. *Acta Crystallogr.*, **1965**, *18*, 1067-1077.

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

SLYNKO, V.; SCHUBERT, M.; NUMAO, S.; KOWARIK, M.; AEBI, M.; ALLAIN, F. H.-T.: NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc*, **2009**, 131, 1274-1281.

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

STICHT, H.; PICKFORD, A. R.; POTTS, J. R.; CAMPBELL, I. D.: Solution Structure of the Glycosylated Second Type 2 Module of Fibronectin. *J. Mol. Biol.*, **1998**, *276*, 177-187.

STUBBS, H. J.; SHIA, M. A.; RICE K. G.: Preparative Purification of Tetraantennary Oligosaccharides from Human Asialyl Orosomucoid. *Analytical biochemistry*, **1997**, 247, 357–365.

SURESHAN, K. M.; MIYASOU, T.; WATANABE, Y.: Crystal structure of 1L-1,2:4,5-di-*O*-isopropylidene-*allo*-inositol; A comparison of its conformation in solid and solution states. *Carbohydr. Res.*, **2004**, *339*, 1551-1555.

TAKAHASHI, M.; YOKOE, S.; ASAHI, M.; LEE, S. H.; LI, W.; OSUMI, D.; MIYOSHI, E.; TANIGUCHI, N.: *N*-glycan of ErbB family plays a crucial role in dimer formation and tumor promotion. *Biochim. Biophys. Acta*, **2008**, *1780*, 520-524.

THANKA CHRISTLET, T. H.; VELURAJA, K.: Database Analysis of O-Glycosylation Sites in Proteins. *Biophys. J.*, **2001**, *80*, 952-960.

The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

THOGERSEN, H.; LEMIEUX, R. U.; BOCK, K.; MEYER, B.: Further Justification for the Exo-Anomeric Effect - Conformational-Analysis Based on Nuclear Magnetic-Resonance Spectroscopy of Oligosaccharides. *Can J. Chem.* **1982**, 60, 44-57.

TRAINOR, G. L.: The importance of plasma protein binding in drug discovery, *Expert Opin. Drug Discov.*, **2007**, *2*, 51-64.

TREUHEIT, M. J.; COSTELLO, C. E.; HALSALL, H. B.: Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein. *Biochem. J.* **1992**, *283*, 105-112.

TURNBULL, J. E.; FIELD, R. A.: Emerging glycomics technologies. *Nat. Chem. Biol.*, **2007**, *3*, 74-77.

VAIDEHI, N., AND KENAKIN, T.: The role of conformational ensembles of seven transmembrane receptors in functional selectivity, *Current opinion in pharmacology*, **2010**, *10*, 775-781.

VAKONAKIS, I.; LANGENHAN, T.; PRÖMEL, S.; RUSS, A.; CAMPBELL, I. D.: Solution Structure and Sugar-Binding Mechanism of Mouse Latrophilin-1 RBL: a 7TM Receptor-Attached Lectin-Like Domain. *Structure*, **2008**, *16*, 944-953.

VAN DER SPOEL, D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A. E.; BERENDSEN, H. J. C.: GROMACS: fast, flexible and free. *J. Comput. Chem.*, **2005**, *26*, 1701-1718.

VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C.: Computer simulations of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, **1990**, *29*, 992-1023.

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, Zurique, Suíça, 1996.

VAN GUNSTEREN, W. F.; MARK, A. E.: Validation of molecular dynamics simulation. *J. Chem. Phys.,* **1998**, *108*, 6109-6116.

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

VERLI, H.; GUIMARÃES, J. A.: Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr. Res.*, **2004**, *339*, 281-290.

VERLI, H.; GUIMARÃES, J. A.: Insights into the induced fit mechanism in antithrombin–heparin interaction using molecular dynamics simulations. *J. Mol. Graph. Model.*, **2005**, *24*, 203-212.

VERLI, H.: *Tese de Doutorado*. Universidade Federal do Rio Grande do Sul, Brasil, **2005**.

VERLI, H.; CALAZANS, A.; BRINDEIRO, R.; TANURI, A.; GUIMARAES, J. A.: Molecular dynamics analysis of HIV-1 matrix protein: clarifying differences between crystallographic and solution structures. *J. Mol. Graph. Mod.*, **2007**, *26*, 62-68.

VESTWEBER, D.; BLANKS, J. E.: Mechanisms that regulate the function of the selectins and their ligands. *Physiol. Rev.*, **1999**, *79*, 181–213.

VIRUDACHALAM, R.; RAO, V. S. R.: Theoretical Studies on Peptidoglycans. II. Conformations of the Disaccharide-Peptide Subunit and the Three-Dimensional Structure of Ptidoglycan. *Biopolymers*, **1979**, *18*, 571-589.

WANG, J.; WOLF, R. M.; CALDWELL, J. W.; KOLLMAN, P. A.; CASE, D. A.: Development and testing of a general amber force field.. *J. Comput. Chem.*, **2004**, *25*, 1157-1174.

WASAN, K. M.; BROCKS, D. R.; LEE, S. D.; SACHS-BARRABLE, K.; THORNTON, S. J.: Impact of lipoproteins on the biological activity and disposition of hydrophobic drugs: implications for drug discovery, *Nature Reviews Drug Discovery* **2008**, *7*, 84-99.

WITHKA, J. M.; WYSS, D. F.; WAGNER, G.; ARULANANDAM, A. R.; REINHERZ, E. L.; RECNY, M. A.: Structure of the glycosylated adhesion domain of human T lymphocyte glycoprotein CD2. *Structure*, **1993**, *1*, 69-81.

WONG, N. K.; RENOUF, D. V.; LEHMANN, S.; HOUNSELL, E. F.: Glycosylation of prions and its effects on protein conformation relevant to amino acid mutations. *J. Mol. Graph. Model.*, **2000**, *18*, 126-134.

WOOD, M. J.; SAMPOLI BENITEZ, B. A.; KOMIVES, E. A.: Solution structure of the smallest cofactor-active fragment of thrombomodulin. *Nat. Struct. Biol.*, **2000**, *7*, 200-204.

WOODS, R. J.; DWEK, R. A.; EDGE, C. J.; FRASER-REID, B.: Molecular Mechanical and Molecular Dynamical Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM-93 Parameter Development. *J. Phys. Chem.*, **1995**, *99*, 3832-3846.

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

WORMALD, M. R.; PETRESCU, A. J.; PAO, Y. -L.; GLITHERO, A.; ELLIOTT, T.; DWEK, R. A.: Conformational Studies of Oligosaccharides and Glycopeptides: Complementarity of NMR, X-ray Crystallography, and Molecular Modelling. *Chem. Rev.*, **2002**, *102*, 371-386.

WYSS, D. F.; WITHKA, J. M.; KNOPPERS, M. H.; STERNE, K. A.; RECNY, M. A.; WAGNER, G.: 1H resonance assignments and secondary structure of the 13.6 kDa glycosylated adhesion domain of human CD2. *Biochemistry*, **1993**, *32*, 10995-11006.

WYSS, D. F.; CHOI, J. S.; LI, J.; KNOPPERS, M. H.; WILLIS, K. J.; ARULANANDAM, A. R. N.; SMOLYAR, A.; REINHERZ, E. L.; WAGNER, G.: Conformation and Function of the N-linked Glycan in the Adhesion Domain of Human CD2. *Science*, **1995**, *269*, 1273-1278.

XIA, J.; DALY, R. P.; CHUANG, F. -C.; PARKER, L.; JENSEN, J. H.; MARGULIS, C. J.: Sugar Folding: A Novel Structural Prediction Tool for Oligosaccharides and Polysaccharides 2. *J. Chem. Theory Comput.*, **2007**, *3*, 1629-1643.

XU, Y.; COLLETIER, J. P.; JIANG, H.; SILMAN, I.; SUSSMAN J. L.; WEIK, M.: Induced-fit or pre-existing equilibrium dynamics? Lessons from protein crystallography and MD simulations on acetylcholinesterase. *Protein Sci.*, **2008**, *17*, 601-605.

YONEDA, J. D.; ALBUQUERQUE, M. G.; LEAL, K. Z.; SEIDL, P. R.; WHEELER, R. A.; BOESCH, S. E.; DE ALENCASTRO, R. B.; DE SOUZA, M. C. B. V.; FERREIRA, V. F.: Molecular dynamics simulations of a nucleoside analogue of 1,4-dihydro-4-oxoquinoline-3-carboxylic acid synthesized as a potential antiviral agent: Conformational studies in vacuum and in water. *Theochem*, **2006**, *778*, 97-103.

ZUEGG, J.; GREADY, J. E.: Molecular dynamics simulation of human prion protein including both N-linked oligosaccharides and the GPI anchor. *Glycobiology*, **2000**, *10*, 959-974.

7 Anexos

Durante a realização desta tese, alguns outros projetos foram elaborados em forma de colaborações com o grupo da professora Márcia Margis-Pinheiro ou orientação de iniciação científica. Destes resultados, surgiram dois trabalhos um aceito (Trabalho I) e outro submetido (Trabalho II), as capas destes trabalhos estão incluídas a seguir.

7.1 Trabalho I

APX-R is a new heme-containing protein functionally associated to APx but evolutionarily divergent

Journal:	*New Phytologist*
Manuscript ID:	NPH-MS-2010-11157.R1
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Lazzarotto, Fernanda; Universidade Federal do Rio Grande do Sul, Genetics Teixeira, Felipe; Universidade Federal do Rio Grande do Sul, Genetics Rosa, Silvia; Universidade Federal do Rio Grande do Sul, Genetics Dunand, Christophe; University of Toulouse, Plant Cell Surfaces and Signaling Laboratory Fernandes, Claudia; Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia Fontenele, Adilton; Universidade Federal do Ceará, Biochemistry Silveira, Albenisio; Universidade Federal do Ceará, Biochemistry Verli, Hugo; Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia Margis, Rogério; Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia Margis-Pinheiro, Marcia; Universidade Federal do Rio Grande do Sul, Genetics
Key Words:	single copy gene, Heme peroxidase, peroxidase, ascorbate peroxidase, APx, antioxidant, H2O2 scavenging

7.2 Trabalho II

Elsevier Editorial System(tm) for Journal of Structural Biology
Manuscript Draft

Manuscript Number:

Title: Structural and conformational characterization of human prothrombin

Article Type: Regular Article

Keywords: Comparative Modeling; Molecular Dynamics; Docking; Coarse-Grained; Coagulation Cascade; Prothrombin.

Corresponding Author: Dr. Hugo Verli, Ph.D.

Corresponding Author's Institution: Federal University of Rio Grande do Sul

First Author: Carla G Chiodi, BSc

Order of Authors: Carla G Chiodi, BSc; Hugo Verli, Ph.D.; Claudia L Fernandes, Ph.D.

Abstract: Prothrombin is an important zymogen of the coagulation cascade. Its active form, α-thrombin, is the key enzyme for converting soluble fibrinogen into insoluble fibrin monomers which, in turn, organize themselves in a proteic clot that sustain the platelet plug and, consequently, contribute to control the hemorrhagic processes. While thrombin has several crystallographic structures deposited on PDB, prothrombin three-dimensional structure is until now absent, impairing the design of new factor Xa inhibitors based on the enzyme-substrate recognition. In this context, the current work employs comparative modelling techniques, docking calculations and molecular dynamics simulations in order to obtain a theoretical model of the complete human prothrombin, capable to support the structural interpretation of its biological roles at the atomic level and so be potentially employed in further efforts for the designing of new antithrombotic agents. While comparative modeling techniques support the building of models for prothrombin between residues 44-187 and 207-622, there is no structural reference for connecting these segments. So models of these regions were built and submitted to docking calculations using Autodock, HEX and ZDOCK and subsequently refined under coarse-grained molecular dynamics simulations within GROMACS. Accordingly, the so obtained data suggest the presence of hinge movements between prothrombin domains within microsecond time scale and, consequently, the co- existence of multiple conformational states in biological solutions. Such flexibility is related to hinge motions between its domains, which may explain the difficulties in obtaining its crystallographic structure.

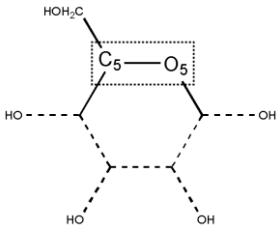
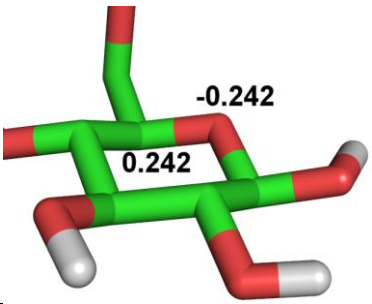
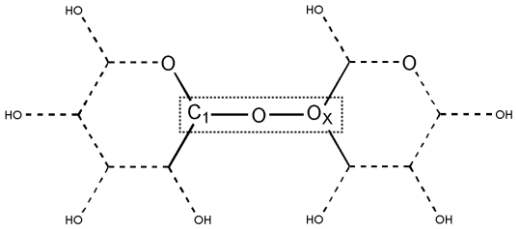
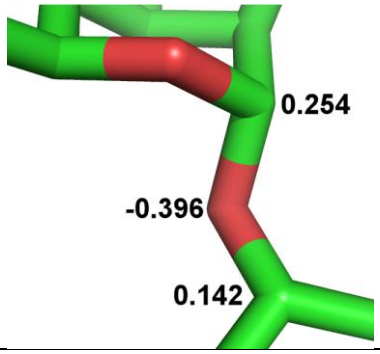
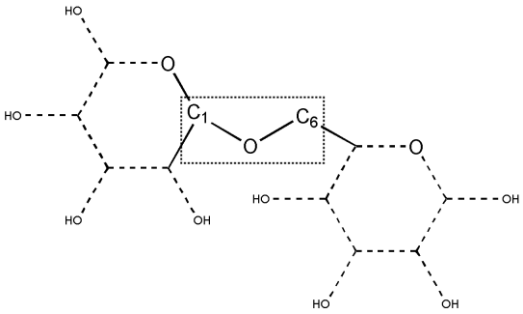
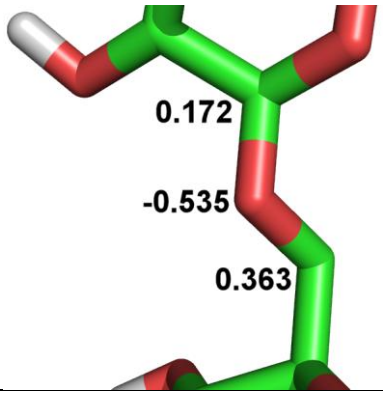
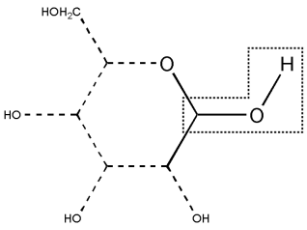
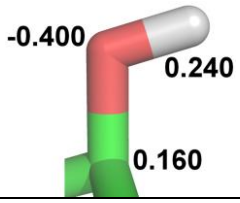
Suggested Reviewers: Richard C Garratt PhD
richard@if.sc.usp.br

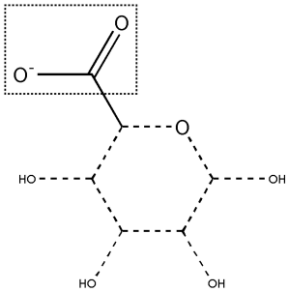
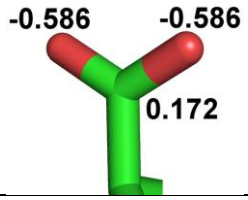
Luca Monticelli PhD
luca.monticelli@inserm.fr

Russolina B Zingali PhD
lzingali@bioqmed.ufrj.br

7.3 Cargas atômicas de Löwdin

As cargas atômicas parciais de Löwdin são divididas em grupamentos, ou grupos de carga, para ser aplicadas para diferentes resíduos de carboidratos visando simulações em diferentes campos de força. As cargas a seguir calculadas (Becker *et al.*, 2007) foram utilizadas nos dissacarídeos estudados.

Grupamento	Estrutura 2D	Estrutura 3D e cargas
C5-O5 do anel piranosídico		
Ligações glicosídicas 1→X*		
Ligação glicosídica 1→6		
Hidroxila		

Grupamento	Estrutura 2D	Estrutura 3D e cargas
Carboxilato		

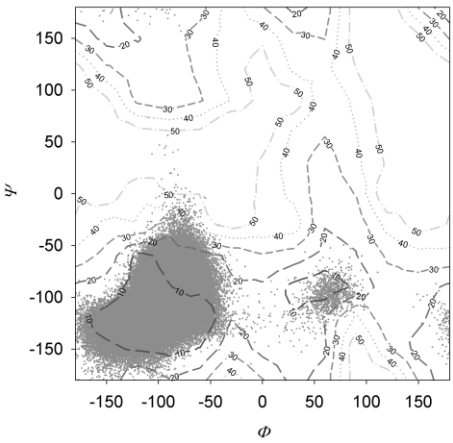
* X= 2, 3 ou 4

8 Apêndices

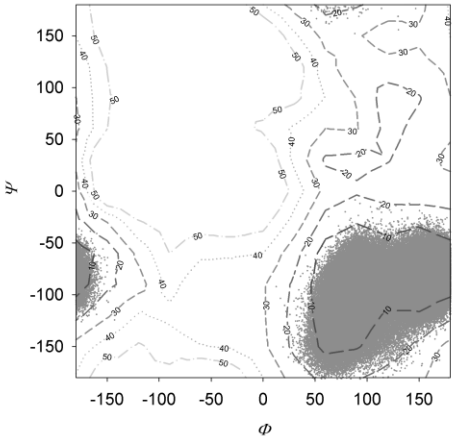
8.1 Mapas de dissacarídeos utilizados nas simulações da AGP

Estes mapas de contorno foram utilizados para calcular as conformações mais estáveis dos dissacarideos que fazem parte da AGP. Ao fundo (cinza) temos os resultados de mínimo de energia para a simulação destes dissacarídeos em solução.

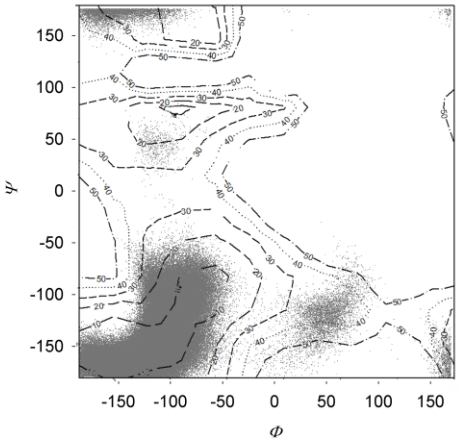
β-GlcNAc-(1→2)-α-Man



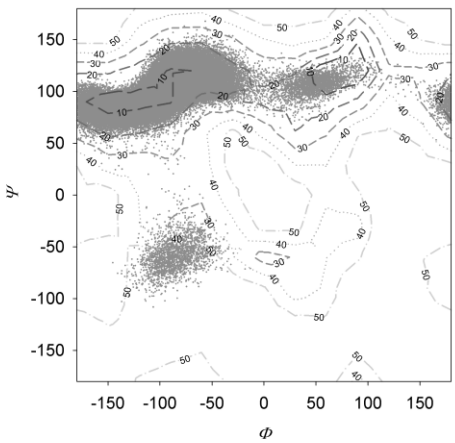
α-Man-(1→3)-β-Man



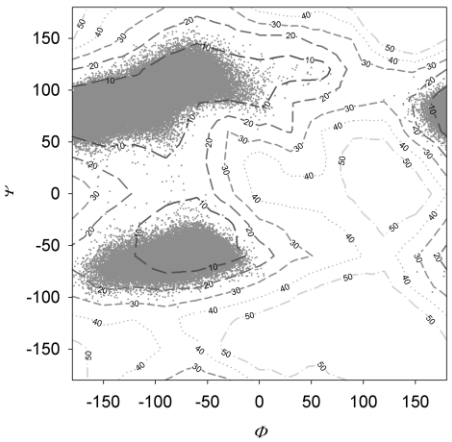
α-Fuc-(1→3)-β-GlcNAc



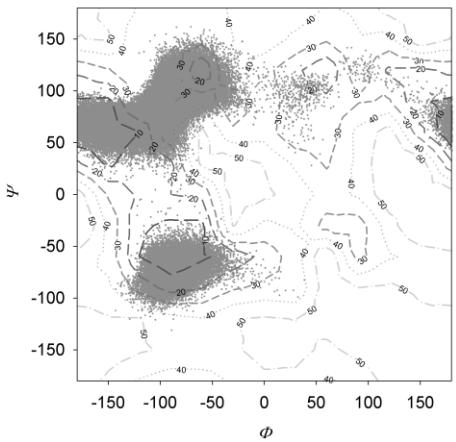
β-Gal-(1→4)-β-GlcNAc

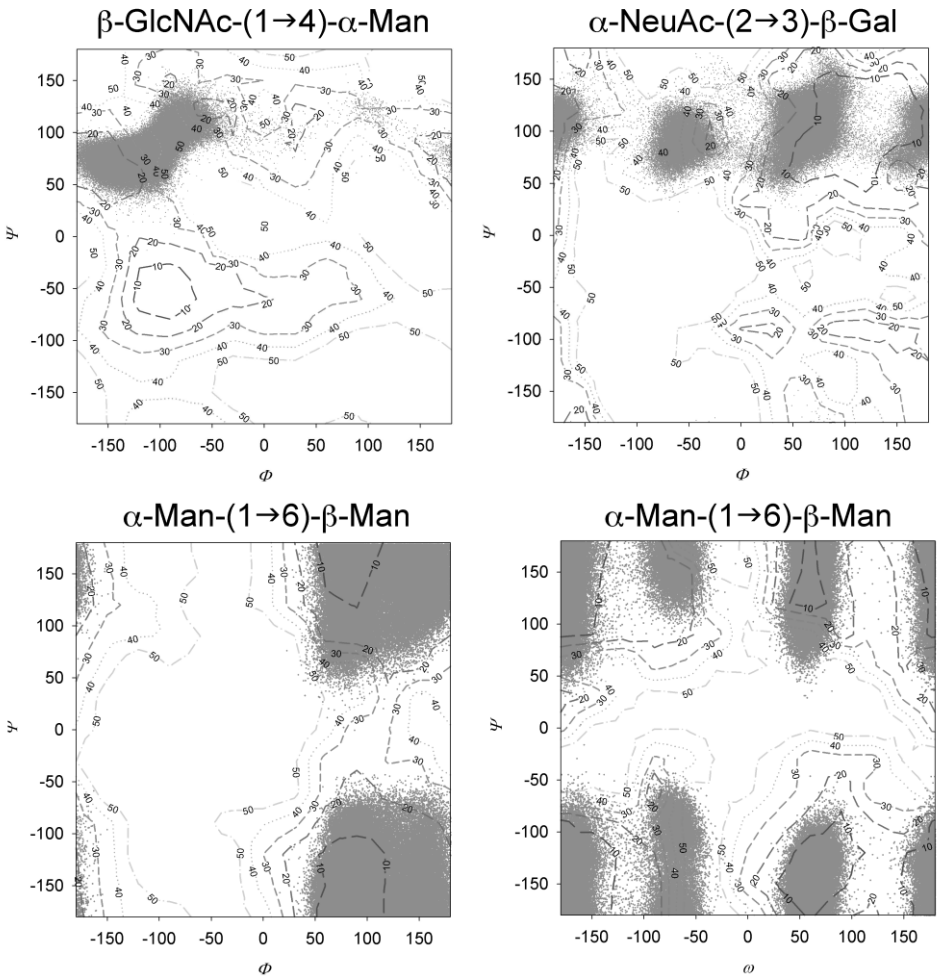


β-Man-(1→4)-β-GlcNAc



β-GlcNAc-(1→4)-β-GlcNAc




β-GlcNAc-(1→4)-α-Man
α-NeuAc-(2→3)-β-Gal
α-Man-(1→6)-β-Man
α-Man-(1→6)-β-Man
ψ
Φ
ω
150
100
50
0
-50
-100
-150

8.2 Topologias de monossacarídeos para GROMOS96 43a1

Esses arquivos contêm os parâmetros de ligação, ângulo de ligação e diedros para as moléculas em estudo. No caso específico do campo de força GROMOS96 43a1, utilizando o pacote GROMACS, essas topologias, juntamente com outras que descrevem outras classes de biomoléculas, fazem parte de um arquivo com extensão “rtp”, localizado em uma pasta de topologias compartilhadas do programa.

Abaixo, seguem as topologias para os monossacarídeos constituintes da glicoproteína AGP, assim como parâmetros para os aminoácidos a que os oligossacarídeos se ligam.

*β-GlcNAc-(1→N)-***Asn**

```
[ ASN2 ]
 [ atoms ]
    N     N    -0.28000     0
    H     H     0.28000     0
   CA   CH1     0.00000     1
   CB   CH2     0.00000     1
   CG     C     0.38000     2
  OD1     O    -0.38000     2
  ND2    NT    -0.26500     3
  HD2     H     0.15500     3
    C     C       0.380     4
    O     O      -0.380     4
 [ bonds ]
    N     H    gb_2
    N    CA    gb_20
   CA     C    gb_26
    C     O    gb_4
    C    +N    gb_9
   CA    CB    gb_26
   CB    CG    gb_26
   CG   OD1    gb_4
   CG   ND2    gb_8
  ND2   HD2    gb_2
 [ angles ]
;  ai    aj    ak   gromos type
   -C     N     H     ga_31
    H     N    CA     ga_17
   -C     N    CA     ga_30
    N    CA     C     ga_12
   CA     C    +N     ga_18
   CA     C     O     ga_29
    O     C    +N     ga_32
   CA     C     O     ga_29
    O     C    +N     ga_32
    N    CA    CB     ga_12
    C    CA    CB     ga_12
   CA    CB    CG     ga_14
   CB    CG   OD1     ga_29
   CB    CG   ND2     ga_18
  OD1    CG   ND2     ga_32
   CG   ND2   HD2     ga_22
 [ impropers ]
;  ai    aj    ak    al   gromos type
    N    -C    CA     H     gi_1
    C    CA    +N     O     gi_1
   CA     N     C    CB     gi_2
   CG   OD1   ND2    CB     gi_1
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -CA    -C     N    CA     gd_4
   -C     N    CA     C     gd_19
    N    CA     C    +N     gd_20
    N    CA    CB    CG     gd_17
   CA    CB    CG   ND2     gd_20
   CB    CG   ND2   HD2     gd_4
```

*β-GlcNAc-(1→4)-**β-GlcNAc**-(1→N)-Asn*

```
[ NAG1 ]
 [ atoms ]
   C8   CH3      0.07000     0
   C7     C      0.27100     0
   O7     O     -0.40500     0
   N2     N     -0.23100     0
  H22     H      0.20100     0
   C2   CH1      0.09400     0
   C1   CH1      0.11000     1
   O5    OA     -0.24200     2
   C5   CH1      0.24200     2
   C6   CH2      0.36300     3
   O6    OA     -0.53500     3
   C4   CH1      0.14200     4
   O4    OA     -0.39600     4
   C3   CH1      0.16000     5
   O3    OA     -0.40000     5
  H32     H      0.24000     5
 [ bonds ]
   C8   C7     gb_26
   C7   O7     gb_4
   C7   N2     gb_10
   N2  H22     gb_2
   N2   C2     gb_20
   C2   C1     gb_25
   C2   C3     gb_25
   C1   O5     gb_19
   O5   C5     gb_19
   C5   C6     gb_25
   C5   C4     gb_25
   C6   O6     gb_19
   C4   O4     gb_19
   C4   C3     gb_25
   O4  +C1     gb_19
   C3   O3     gb_19
   O3  H32     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C8   C7   O7       ga_29
   C8   C7   N2       ga_18
   O7   C7   N2       ga_32
   C7   N2  H22       ga_31
   C7   N2   C2       ga_30
  H22   N2   C2       ga_24
   N2   C2   C1       ga_12
   N2   C2   C3       ga_12
   C1   C2   C3       ga_7
   C2   C1   O5       ga_8
   C1   O5   C5       ga_9
   O5   C5   C6       ga_8
   O5   C5   C4       ga_8
   C6   C5   C4       ga_7
   C5   C6   O6       ga_8
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   O4   C4   C3       ga_8
   C4   O4  +C1       ga_9
   C2   C3   C4       ga_7
   C2   C3   O3       ga_8
   C4   C3   O3       ga_8
   C3   O3  H32       ga_11
[ impropers ]
;  ai    aj    ak    al   gromos type
   C7    C8    O7    N2     gi_1
   N2    C7   H22    C2     gi_1
   C2    N2    C3    C1     gi_2
   C5    O5    C6    C4     gi_2
   C4    C5    C3    O4     gi_2
   C3    O3    C2    C4     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C8    C7    N2    C2     gd_4
   C3    C2    N2    C7     gd_14
   O5    C1    C2    N2     gd_8
   O3    C3    C2    N2     gd_8
   C2    C1    O5    C5     gd_14
   C4    C5    O5    C1     gd_14
   O6    C6    C5    O5     gd_8
   C3    C4    C5    O5     gd_7
   O3    C3    C4    C5     gd_7
   C2    C3    O3   H32     gd_12
   C3    C4    O4   +C1     gd_14
   C2    C3    C4    C5     gd_17
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   N2    C2    C3    C4     gd_7
   C1    C2    C3    O3     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C1    C2    C3     gd_7
   C1    C2    C3    C4     gd_17
   C2    C3    C4    O4     gd_7
```

*β-Man-(1→4)-***β-GlcNAc***-(1→4)-β-GlcNAc*

```
[ NAG2 ]
 [ atoms ]
   C8   CH3      0.07000     0
   C7     C      0.27100     0
   O7     O     -0.40500     0
   N2     N     -0.23100     0
  H22     H      0.20100     0
   C2   CH1      0.09400     0
   C1   CH1      0.25400     1
   O5    OA     -0.24200     2
   C5   CH1      0.24200     2
   C6   CH2      0.16000     3
   O6    OA     -0.40000     3
  H63     H      0.24000     3
   C4   CH1      0.14200     4
   O4    OA     -0.39600     4
   C3   CH1      0.16000     5
   O3    OA     -0.40000     5
  H32     H      0.24000     5
 [ bonds ]
   C8   C7     gb_26
   C7   O7     gb_4
   C7   N2     gb_10
   N2  H22     gb_2
   N2   C2     gb_20
   C2   C1     gb_25
   C2   C3     gb_25
   C1   O5     gb_19
   O5   C5     gb_19
   C5   C6     gb_25
   C5   C4     gb_25
   C6   O6     gb_19
   O6  H63     gb_1
   C4   O4     gb_19
   O4  +C1     gb_19
   C4   C3     gb_25
   C3   O3     gb_19
   O3  H32     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C8   C7   O7       ga_29
   C8   C7   N2       ga_18
   O7   C7   N2       ga_32
   C7   N2  H22       ga_31
   C7   N2   C2       ga_30
  H22   N2   C2       ga_24
   N2   C2   C1       ga_12
   N2   C2   C3       ga_12
   C1   C2   C3       ga_7
   C2   C1   O5       ga_8
   C1   O5   C5       ga_9
  -O4   C1   O5       ga_8
  -O4   C1   C2       ga_8
   O5   C5   C6       ga_8
   O5   C5   C4       ga_8
   C6   C5   C4       ga_7
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   C4   O4  +C1       ga_9
   O4   C4   C3       ga_8
   C2   C3   C4       ga_7
   C2   C3   O3       ga_8
   C4   C3   O3       ga_8
   C3   O3  H32       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O4    O5    C2     gi_2
   C7    C8    O7    N2     gi_1
   N2    C7   H22    C2     gi_1
   C2    N2    C3    C1     gi_2
   C5    O5    C6    C4     gi_2
   C4    C5    C3    O4     gi_2
   C3    O3    C2    C4     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C8    C7    N2    C2     gd_4
   C3    C2    N2    C7     gd_14
   O5    C1    C2    N2     gd_8
   O3    C3    C2    N2     gd_8
   C2    C1    O5    C5     gd_14
   C4    C5    O5    C1     gd_14
  -C4   -O4    C1    C2     gd_14
  -O4    C1    C2    C3     gd_17
  -O4    C1    C2    C3     gd_7
  -O4    C1    C2    O2     gd_8
   O6    C6    C5    O5     gd_8
   C3    C4    C5    O5     gd_7
   O3    C3    C4    C5     gd_7
   C2    C3    O3   H32     gd_12
   C3    C4    O4   +C1     gd_14
  H63    O6    C6    C5     gd_12
   C2    C3    C4    C5     gd_17
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   N2    C2    C3    C4     gd_7
   C1    C2    C3    O3     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C1    C2    C3     gd_7
   C1    C2    C3    C4     gd_17
   C2    C3    C4    O4     gd_7
```

*α-Man-(1→6)-***β-Man***-(1→4)-β-GlcNAc*
3
↑
1
α-Man

```
[ BMA1 ]
 [ atoms ]
   C6   CH2     0.36300     0
   O6    OA    -0.53500     0
         C5   CH1     0.24200
      1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.16000     3
   O2    OA    -0.40000     3
  H22     H     0.24000     3
   C3   CH1     0.14200     4
   O3    OA    -0.39600     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   O2  H22     gb_1
   C3   O3     gb_19
   C3   C4     gb_25
   C4   O4     gb_19
   O4  H42     gb_1

1.1.1.1.1      C6   O6     gb_19
   O6  +C1     gb_19
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
  -O4   C1   O5       ga_8
  -O4   C1   C2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   O2  H22       ga_11
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
   C5   C6   O6       ga_8
   C6   O6  +C1       ga_9
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O4    O5    C2     gi_2
   C5    C6    C4    O5     gi_2
   C2    C3    O2    C1     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C1    C2    O2   H22     gd_12
   C4    C3    C2    C1     gd_17
   O4    C4    C3    C2     gd_7
   C5    C4    O4   H42     gd_12
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
  -C4   -O4    C1    C2     gd_14
  -O4    C1    C2    C3     gd_17
  -O4    C1    C2    C3     gd_7
  -O4    C1    C2    O2     gd_8
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   C5    C6    O6   +C1     gd_14
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
```

β-GlcNAc-*(1→2)*-**α-Man**-*(1→6)*-*β-Man*

```
[ MAN1 ]
 [ atoms ]
   C6   CH2     0.16000     0
   O6    OA    -0.40000     0
  H63     H     0.24000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.17200     2
   C2   CH1     0.14200     3
   O2    OA    -0.39600     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   C3   O3     gb_19
   O3  H32     gb_1
   C3   C4     gb_25
   C4   O4     gb_19
   O4  H42     gb_1
   C6   O6     gb_19
   O6  H63     gb_1
   O2  +C1     gb_19
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
  -O6   C1   O5       ga_8
  -O6   C1   C2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   C3   O3       ga_8
   C3   O3  H32       ga_11
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C2   O2  +C1       ga_9
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O6    C2    O5     gi_2
   C5    C6    C4    O5     gi_2
   C2    C3    O2    C1     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C4    C3    C2    C1     gd_17
   O4    C4    C3    C2     gd_7
   C5    C4    O4   H42     gd_12
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
  -C6   -O6    C1    C2     gd_14
  -O6    C1    C2    C3     gd_17
  -O6    C1    C2    C3     gd_7
  -O6    C1    C2    O2     gd_8
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   C3    C2    O2   +C1     gd_14
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C1    C2    C3    O3     gd_7
   C2    C3    O3   H32     gd_12
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   H63     gd_12
```

*β-GlcNAc-(1→2)-***α-Man***-(1→6)-β-Man*
4
↑
1
β-GlcNAc

```
[ MAN5 ]
 [ atoms ]
   C6   CH2     0.16000     0
   O6    OA    -0.40000     0
  H63     H     0.24000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.17200     2
   C2   CH1     0.14200     3
   O2    OA    -0.39600     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.14200     5
   O4    OA    -0.39600     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   C3   O3     gb_19
   O3  H32     gb_1
   C3   C4     gb_25
   C4   O4     gb_19
   C6   O6     gb_19
   O6  H63     gb_1
   O2  +C1     gb_19
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
  -O6   C1   O5       ga_8
  -O6   C1   C2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   C3   O3       ga_8
   C3   O3  H32       ga_11
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C2   O2  +C1       ga_9
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O6    C2    O5     gi_2
   C5    C6    C4    O5     gi_2
   C2    C3    O2    C1     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C4    C3    C2    C1     gd_17
   O4    C4    C3    C2     gd_7
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
  -C6   -O6    C1    C2     gd_14
  -O6    C1    C2    C3     gd_17
  -O6    C1    C2    C3     gd_7
  -O6    C1    C2    O2     gd_8
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   C3    C2    O2   +C1     gd_14
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C1    C2    C3    O3     gd_7
   C2    C3    O3   H32     gd_12
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   H63     gd_12
```

β-GlcNAc-*(1→2)*-**α-Man**-*(1→3)*-*β-Man*

```
[ MAN8 ]
 [ atoms ]
   C6   CH2     0.16000     0
   O6    OA    -0.40000     0
  H63     H     0.24000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.14200     3
   O2    OA    -0.39600     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   O2  +C1     gb_19
   C2   C3     gb_25
   C3   O3     gb_19
   C3   C4     gb_25
   O3  H32     gb_1
   C4   O4     gb_19
   O4  H42     gb_1
   C6   O6     gb_19
   O6  H63     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   O2  +C1       ga_9
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C3   O3  H32       ga_11
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C5    C6    C4    O5     gi_2
   C2    C3    O2    C1     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C4    C3    C2    C1     gd_17
   C2    C3    O3   H32     gd_12
   O4    C4    C3    C2     gd_7
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   C3    C4    O4   H42     gd_12
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C3    C2    O2   +C1     gd_14
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   H63     gd_12
```

*β-GlcNAc-(1→2)-***α-Man***-(1→3)-β-Man*
4
↑
1
β-GlcNAc

```
[ MAN7 ]
 [ atoms ]
   C6   CH2     0.16000     0
   O6    OA    -0.40000     0
  H63     H     0.24000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.14200     3
   O2    OA    -0.39600     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.14200     5
   O4    OA    -0.39600     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   O2  +C1     gb_19
   C2   C3     gb_25
   C3   O3     gb_19
   C3   C4     gb_25
   O3  H32     gb_1
   C4   O4     gb_19
   C6   O6     gb_19
   O6  H63     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   O2  +C1       ga_9
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C3   O3  H32       ga_11
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C5    C6    C4    O5     gi_2
   C2    C3    O2    C1     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C4    C3    C2    C1     gd_17
   C2    C3    O3   H32     gd_12
   O4    C4    C3    C2     gd_7
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C3    C2    O2   +C1     gd_14
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   H63     gd_12
```

*β-Gal-(1→4)-***β-GlcNAc***-(1→2)-α-Man*

```
[ NAG9 ]
 [ atoms ]
   C8   CH3     0.07000     0
   C7     C     0.27100     0
   O7     O    -0.40500     0
   N2     N    -0.23100     0
  H22     H     0.20100     0
   C2   CH1     0.09400     0
   C1   CH1     0.25400     1
   O5    OA    -0.24200     2
   C5   CH1     0.24200     2
   C6   CH2     0.16000     3
   O6    OA    -0.40000     3
  H63     H     0.24000     3
   C4   CH1     0.14200     4
   O4    OA    -0.39600     4
   C3   CH1     0.16000     5
   O3    OA    -0.40000     5
  H32     H     0.24000     5
 [ bonds ]
   C8   C7     gb_26
   C7   O7     gb_4
   C7   N2     gb_10
   N2  H22     gb_2
   N2   C2     gb_20
   C2   C1     gb_25
   C2   C3     gb_25
   C1   O5     gb_19
   O5   C5     gb_19
   C5   C6     gb_25
   C5   C4     gb_25
   C6   O6     gb_19
   O6  H63     gb_1
   C4   O4     gb_19
   O4  +C1     gb_19
   C4   C3     gb_25
   C3   O3     gb_19
   O3  H32     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C8   C7   O7       ga_29
   C8   C7   N2       ga_18
   O7   C7   N2       ga_32
   C7   N2  H22       ga_31
   C7   N2   C2       ga_30
  H22   N2   C2       ga_24
   N2   C2   C1       ga_12
   N2   C2   C3       ga_12
   C1   C2   C3       ga_7
   C2   C1   O5       ga_8
   C1   O5   C5       ga_9
  -O2   C1   O5       ga_8
  -O2   C1   C2       ga_8
   O5   C5   C6       ga_8
   O5   C5   C4       ga_8
   C6   C5   C4       ga_7
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   C4   O4  +C1       ga_9
   O4   C4   C3       ga_8
   C2   C3   C4       ga_7
   C2   C3   O3       ga_8
   C4   C3   O3       ga_8
   C3   O3  H32       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O2    O5    C2     gi_2
   C7    C8    O7    N2     gi_1
   N2    C7   H22    C2     gi_1
   C2    N2    C3    C1     gi_2
   C5    O5    C6    C4     gi_2
   C4    C5    C3    O4     gi_2
   C3    O3    C2    C4     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C8    C7    N2    C2     gd_4
   C3    C2    N2    C7     gd_14
   O5    C1    C2    N2     gd_8
   O3    C3    C2    N2     gd_8
   C2    C1    O5    C5     gd_14
   C4    C5    O5    C1     gd_14
  -C2   -O2    C1    C2     gd_14
  -O2    C1    C2    C3     gd_17
  -O2    C1    C2    C3     gd_7
  -O2    C1    C2    O2     gd_8
   O6    C6    C5    O5     gd_8
   C3    C4    C5    O5     gd_7
   O3    C3    C4    C5     gd_7
   C2    C3    O3   H32     gd_12
   C3    C4    O4   +C1     gd_14
  H63    O6    C6    C5     gd_12
   C2    C3    C4    C5     gd_17
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   N2    C2    C3    C4     gd_7
   C1    C2    C3    O3     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C1    C2    C3     gd_7
   C1    C2    C3    C4     gd_17
   C2    C3    C4    O4     gd_7
```

*β-Gal-(1→4)-***β-GlcNAc***-(1→4)-α-Man*

```
[ NAG7 ]
 [ atoms ]
   C8   CH3     0.07000     0
   C7     C     0.27100     0
   O7     O    -0.40500     0
   N2     N    -0.23100     0
  H22     H     0.20100     0
   C2   CH1     0.09400     0
   C1   CH1     0.25400     1
   O5    OA    -0.24200     2
   C5   CH1     0.24200     2
   C6   CH2     0.16000     3
   O6    OA    -0.40000     3
  H63     H     0.24000     3
   C4   CH1     0.14200     4
   O4    OA    -0.39600     4
   C3   CH1     0.16000     5
   O3    OA    -0.40000     5
  H32     H     0.24000     5
 [ bonds ]
   C8   C7     gb_26
   C7   O7     gb_4
   C7   N2     gb_10
   N2  H22     gb_2
   N2   C2     gb_20
   C2   C1     gb_25
   C2   C3     gb_25
   C1   O5     gb_19
   O5   C5     gb_19
   C5   C6     gb_25
   C5   C4     gb_25
   C6   O6     gb_19
   O6  H63     gb_1
   C4   O4     gb_19
   O4  +C1     gb_19
   C4   C3     gb_25
   C3   O3     gb_19
   O3  H32     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C8   C7   O7       ga_29
   C8   C7   N2       ga_18
   O7   C7   N2       ga_32
   C7   N2  H22       ga_31
   C7   N2   C2       ga_30
  H22   N2   C2       ga_24
   N2   C2   C1       ga_12
   N2   C2   C3       ga_12
   C1   C2   C3       ga_7
   C2   C1   O5       ga_8
   C1   O5   C5       ga_9
   O5   C5   C6       ga_8
   O5   C5   C4       ga_8
   C6   C5   C4       ga_7
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   C4   O4  +C1       ga_9
   O4   C4   C3       ga_8
   C2   C3   C4       ga_7
   C2   C3   O3       ga_8
   C4   C3   O3       ga_8
   C3   O3  H32       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C7    C8    O7    N2     gi_1
   N2    C7   H22    C2     gi_1
   C2    N2    C3    C1     gi_2
   C5    O5    C6    C4     gi_2
   C4    C5    C3    O4     gi_2
   C3    O3    C2    C4     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C8    C7    N2    C2     gd_4
   C3    C2    N2    C7     gd_14
   O5    C1    C2    N2     gd_8
   O3    C3    C2    N2     gd_8
   C2    C1    O5    C5     gd_14
   C4    C5    O5    C1     gd_14
   O6    C6    C5    O5     gd_8
   C3    C4    C5    O5     gd_7
   O3    C3    C4    C5     gd_7
   C2    C3    O3   H32     gd_12
   C3    C4    O4   +C1     gd_14
  H63    O6    C6    C5     gd_12
   C2    C3    C4    C5     gd_17
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   N2    C2    C3    C4     gd_7
   C1    C2    C3    O3     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C1    C2    C3     gd_7
   C1    C2    C3    C4     gd_17
   C2    C3    C4    O4     gd_7
```

*β-Gal-(1→4)-***β-GlcNAc***-(1→4)-α-Man*
3
↑
1
α-Fuc

```
[ NAG8 ]
 [ atoms ]
   C8   CH3      0.07000     0
   C7     C      0.27100     0
   O7     O     -0.40500     0
   N2     N     -0.23100     0
  H22     H      0.20100     0
   C2   CH1      0.09400     0
   C1   CH1      0.25400     1
   O5    OA     -0.24200     2
   C5   CH1      0.24200     2
   C6   CH2      0.16000     3
   O6    OA     -0.40000     3
  H63     H      0.24000     3
   C4   CH1      0.14200     4
   O4    OA     -0.39600     4
   C3   CH1      0.14200     5
   O3    OA     -0.39600     5
 [ bonds ]
   C8   C7     gb_26
   C7   O7     gb_4
   C7   N2     gb_10
   N2  H22     gb_2
   N2   C2     gb_20
   C2   C1     gb_25
   C2   C3     gb_25
   C1   O5     gb_19
   O5   C5     gb_19
   C5   C6     gb_25
   C5   C4     gb_25
   C6   O6     gb_19
   O6  H63     gb_1
   C4   O4     gb_19
   O4  +C1     gb_19
   C4   C3     gb_25
   C3   O3     gb_19
 [ angles ]
;  ai    aj    ak   gromos type
   C8   C7   O7       ga_29
   C8   C7   N2       ga_18
   O7   C7   N2       ga_32
   C7   N2  H22       ga_31
   C7   N2   C2       ga_30
  H22   N2   C2       ga_24
   N2   C2   C1       ga_12
   N2   C2   C3       ga_12
   C1   C2   C3       ga_7
   C2   C1   O5       ga_8
   C1   O5   C5       ga_9
   O5   C5   C6       ga_8
   O5   C5   C4       ga_8
   C6   C5   C4       ga_7
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   C4   O4  +C1       ga_9
   O4   C4   C3       ga_8
   C2   C3   C4       ga_7
   C2   C3   O3       ga_8
   C4   C3   O3       ga_8
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C7    C8    O7    N2     gi_1
   N2    C7   H22    C2     gi_1
   C2    N2    C3    C1     gi_2
   C5    O5    C6    C4     gi_2
   C4    C5    C3    O4     gi_2
   C3    O3    C2    C4     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C8    C7    N2    C2     gd_4
   C3    C2    N2    C7     gd_14
   O5    C1    C2    N2     gd_8
   O3    C3    C2    N2     gd_8
   C2    C1    O5    C5     gd_14
   C4    C5    O5    C1     gd_14
   O6    C6    C5    O5     gd_8
   C3    C4    C5    O5     gd_7
   O3    C3    C4    C5     gd_7
   C3    C4    O4   +C1     gd_14
  H63    O6    C6    C5     gd_12
   C2    C3    C4    C5     gd_17
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   N2    C2    C3    C4     gd_7
   C1    C2    C3    O3     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C1    C2    C3     gd_7
   C1    C2    C3    C4     gd_17
   C2    C3    C4    O4     gd_7
```

β-Gal-*(1→4)-β-GlcNAc*

```
[ GAL1 ]
 [ atoms ]
   C6   CH2     0.16000     0
   O6    OA    -0.40000     0
  H63     H     0.24000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.16000     3
   O2    OA    -0.40000     3
  H22     H     0.24000     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   O2  H22     gb_1
   C3   O3     gb_19
   C3   C4     gb_25
   O3  H32     gb_1
   C4   O4     gb_19
   O4  H42     gb_1
   C6   O6     gb_19
   O6  H63     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
  -O4   C1   O5       ga_8
  -O4   C1   C2       ga_8
   C2   O2  H22       ga_11
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C3   O3  H32       ga_11
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
   C5   C6   O6       ga_8
   C6   O6  H63       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O4    O5    C2     gi_2
         C5    C6    C4    O5     gi_2
   C2    C3    C1    O2     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C3    C5     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C1    C2    O2   H22     gd_12
   C4    C3    C2    C1     gd_17
   C2    C3    O3   H32     gd_12
   O4    C4    C3    C2     gd_7
   C5    C4    O4   H42     gd_12
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
  -C4   -O4    C1    C2     gd_14
  -O4    C1    C2    C3     gd_17
  -O4    C1    C2    C3     gd_7
  -O4    C1    C2    O2     gd_8
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   H63     gd_12
```

*αNeuAc-(2→3)-***β-Gal**-*(1→4)-β-GlcNAc*

```
[ GAL3 ]
 [ atoms ]
   C6   CH2     0.36300     0
   O6    OA    -0.53500     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.16000     3
   O2    OA    -0.40000     3
  H22     H     0.24000     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   O2  H22     gb_1
   C3   O3     gb_19
   C3   C4     gb_25
   O3  H32     gb_1
   C4   O4     gb_19
   O4  H42     gb_1
   C6   O6     gb_19
   O6  +C2     gb_19
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
  -O4   C1   O5       ga_8
  -O4   C1   C2       ga_8
   C2   O2  H22       ga_11
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C3   O3  H32       ga_11
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
   C5   C6   O6       ga_8
   C6   O6  +C2       ga_9
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C1   -O4    O5    C2     gi_2
   C5    C6    C4    O5     gi_2
   C2    C3    C1    O2     gi_2
   C3    C4    O3    C2     gi_2
   C4    O4    C3    C5     gi_2
   C5    C2    C4    C1     gi_7
   C5    C2    C3    C1     gi_8
   C5    C2    C3    O5     gi_9
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C1    C2    O2   H22     gd_12
   C4    C3    C2    C1     gd_17
   C2    C3    O3   H32     gd_12
   O4    C4    C3    C2     gd_7
   C5    C4    O4   H42     gd_12
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   C6    C5    C4    O4     gd_7
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
  -C4   -O4    C1    C2     gd_14
  -O4    C1    C2    C3     gd_17
  -O4    C1    C2    C3     gd_7
  -O4    C1    C2    O2     gd_8
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
   C4    C5    C6    O6     gd_7
   C4    C5    C6    O6     gd_17
   O5    C5    C6    O6     gd_8
   C5    C6    O6   +C2     gd_14
```

α-Fuc-*(1→3)-β-GlcNAc*

```
[ FUC1 ]
 [ atoms ]
   C6   CH3     0.00000     0
   C5   CH1     0.24200     1
   O5    OA    -0.24200     1
   C1   CH1     0.25400     2
   C2   CH1     0.16000     3
   O2    OA    -0.40000     3
  H22     H     0.24000     3
   C3   CH1     0.16000     4
   O3    OA    -0.40000     4
  H32     H     0.24000     4
   C4   CH1     0.16000     5
   O4    OA    -0.40000     5
  H42     H     0.24000     5
 [ bonds ]
   C6   C5     gb_25
   C5   O5     gb_19
   C5   C4     gb_25
   O5   C1     gb_19
   C1   C2     gb_25
   C2   O2     gb_19
   C2   C3     gb_25
   O2  H22     gb_1
   C3   O3     gb_19
   C3   C4     gb_25
   O3  H32     gb_1
   C4   O4     gb_19
   O4  H42     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
   C6   C5   O5       ga_8
   C6   C5   C4       ga_7
   O5   C5   C4       ga_8
   C5   O5   C1       ga_9
   O5   C1   C2       ga_8
   C1   C2   O2       ga_8
   C1   C2   C3       ga_7
   O2   C2   C3       ga_8
   C2   O2  H22       ga_48
   C2   C3   O3       ga_8
   C2   C3   C4       ga_7
   O3   C3   C4       ga_8
   C3   O3  H32       ga_11
   C5   C4   C3       ga_7
   C5   C4   O4       ga_8
   C3   C4   O4       ga_8
   C4   O4  H42       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
   C5    C6    O5    C4     gi_2
   C2    C1    C3    O2     gi_2
   C3    C2    O3    C4     gi_2
   C4    O4    C5    C3     gi_2
   C5    C2    C4    C1     gi_4
   C5    C2    C3    C1     gi_5
   C5    C2    C3    O5     gi_6
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
   O4    C4    C5    C6     gd_7
   C2    C1    O5    C5     gd_14
   C3    C2    C1    O5     gd_7
   C1    C2    O2   H22     gd_12
   C4    C3    C2    C1     gd_17
   C2    C3    O3   H32     gd_12
   O4    C4    C3    C2     gd_7
   C5    C4    O4   H42     gd_12
   C2    C3    C4    C5     gd_17
   C1    O5    C5    C4     gd_14
   O5    C5    C4    C3     gd_7
   C6    C5    C4    C3     gd_17
   O5    C1    C2    O2     gd_8
   O2    C2    C3    C4     gd_7
   O2    C2    C3    O3     gd_8
   C1    C2    C3    O3     gd_7
   O3    C3    C4    C5     gd_7
   O3    C3    C4    O4     gd_8
   O5    C5    C4    O4     gd_8
```

α-NeuAc-*(2→3)-β-Gal*

```
[ SIA1 ]
 [ atoms ]
  C11   CH3     0.07000     0
  C10     C     0.27100     0
  O10     O    -0.40500     0
   N5     N    -0.23100     0
  H52     H     0.20100     0
   C5   CH1     0.09400     0
   C4   CH1     0.16000     1
   O4    OA    -0.40000     1
  H42     H     0.24000     1
   C3   CH1     0.00000     2
   C2   CH1     0.25400     3
   C1     C     0.17200     4
  O1B    OM    -0.58600     4
  O1A    OM    -0.58600     4
   O6    OA    -0.24200     5
   C6   CH1     0.24200     5
   C7   CH1     0.16000     6
   O7    OA    -0.40000     6
  H73     H     0.24000     6
   C8   CH1     0.16000     7
   O8    OA    -0.40000     7
  H83     H     0.24000     7
   C9   CH2     0.16000     8
   O9    OA    -0.40000     8
  H93     H     0.24000     8
 [ bonds ]
  C10  C11     gb_26
  C10  O10     gb_4
  C10   N5     gb_10
   N5  H52     gb_2
   C5   N5     gb_20
   C5   C4     gb_25
   C5   C6     gb_25
   C4   O4     gb_19
   C4   C3     gb_25
   O4  H42     gb_1
   C2   C3     gb_25
   C2   C1     gb_25
   C2   O6     gb_19
   C1  O1B     gb_19
   C1  O1A     gb_19
   C6   O6     gb_19
   C6   C7     gb_25
   C7   O7     gb_19
   C7   C8     gb_25
   O7  H73     gb_1
   C8   O8     gb_19
   C8   C9     gb_25
   O8  H83     gb_1
   C9   O9     gb_19
   O9  H93     gb_1
 [ angles ]
;  ai    aj    ak   gromos type
  -O3   C2   O6       ga_8
  -O3   C2   C3       ga_8
  -O3   C2   C1       ga_8
  C11  C10  O10       ga_29
  C11  C10   N5       ga_18
  O10  C10   N5       ga_32
  C10   N5  H52       ga_31
  C10   N5   C5       ga_30
  H52   N5   C5       ga_24
   N5   C5   C4       ga_12
   N5   C5   C6       ga_12
   C4   C5   C6       ga_7
   C5   C4   O4       ga_8
   C5   C4   C3       ga_7
   O4   C4   C3       ga_8
   C4   O4  H42       ga_11
   C4   C3   C2       ga_7
   C3   C2   C1       ga_7
   C3   C2   O6       ga_8

   C1   C2   O6       ga_8
   C2   C1  O1B       ga_8
   C2   C1  O1A       ga_8
  O1B   C1  O1A       ga_37
   C2   O6   C6       ga_9
   C5   C6   O6       ga_8
   C5   C6   C7       ga_7
   O6   C6   C7       ga_8
   C6   C7   O7       ga_8
   C6   C7   C8       ga_7
   O7   C7   C8       ga_8
   C7   O7  H73       ga_11
   C7   C8   O8       ga_8
   C7   C8   C9       ga_7
   O8   C8   C9       ga_8
   C8   O8  H83       ga_11
   C8   C9   O9       ga_8
   C9   O9  H93       ga_11
 [ impropers ]
;  ai    aj    ak    al   gromos type
  C10   C11   O10    N5     gi_1
   N5   C10   H52    C5     gi_1
   C5    N5    C4    C6     gi_2
   C4    C5    C3    O4     gi_2
   C2    C3    C1    O6     gi_2
   C2   -O3    O6    C1     gi_2
   C1    C2   O1B   O1A     gi_1
   C6    C7    O6    C5     gi_2
   C7    C6    O7    C8     gi_2
   C8    C7    O8    C9     gi_2
   C6    C3    C5    C2     gi_4
   C6    C3    C4    C2     gi_5
   C6    C3    C4    O6     gi_6
 [ dihedrals ]
;  ai    aj    ak    al   gromos type
  -C3   -O3    C2    C3     gd_14
  -O3    C2    C3    C4     gd_17
  -O3    C2    C3    C4     gd_7
  -O3    C2    C3    O3     gd_8
  C11   C10    N5    C5     gd_4
```

C6	C5	N5	C10	gd_14
C3	C4	C5	N5	gd_8
C7	C6	C5	N5	gd_8
C5	C4	O4	H42	gd_12
C2	C3	C4	C5	gd_17
O6	C2	C3	C4	gd_7
C3	C2	C1	O1A	gd_20
C3	C2	C1	O1B	gd_20
C3	C2	O6	C6	gd_14
C3	C4	C5	C6	gd_17
C2	O6	C6	C5	gd_14
O6	C6	C5	C4	gd_7
C7	C6	C5	C4	gd_17
O6	C2	C3	O3	gd_8
O3	C3	C4	C5	gd_7
O3	C3	C4	O4	gd_8
C2	C3	C4	O4	gd_7
O4	C4	C5	C6	gd_7
O4	C4	C5	N5	gd_8
O6	C6	C5	N5	gd_8
C7	C6	O6	C2	gd_14
C8	C7	C6	C5	gd_17
O7	C7	C6	C5	gd_7
O7	C7	C8	C9	gd_7
O7	C7	C6	O6	gd_8
C6	C7	O7	H73	gd_12
C9	C8	C7	C6	gd_17
O8	C8	C7	C6	gd_7
O8	C8	C7	O7	gd_8
C7	C8	O8	H83	gd_12
O9	C9	C8	O8	gd_8
O9	C9	C8	C7	gd_7
C8	C9	O9	H93	gd_12

8.3 Tabelas de parâmetros de ligação, ângulos e diedros próprios e impróprios para GROMOS96 43a1

Esses arquivos contêm os parâmetros utilizados em cada ligação, ângulo de ligação e diedros descritos nas topologias citadas anteriormente no campo de força GROMOS96 43a1, utilizando o pacote GROMACS durante as simulações realizadas no decorrer destes trabalhos.

```
;       GROMOS bond-stretching parameters
;
;ICB(H)[N]    CB[N] B0[N]
;
#define gb_1        0.1000  1.5700e+07
; H  -  OA      750
;
#define gb_2        0.1000  1.8700e+07
; H  -  N (all) 895
;
#define gb_3        0.1090  1.2300e+07
; HC  -  C      700
;
#define gb_4        0.1230  1.6600e+07
; C  - O        1200
;
#define gb_5        0.1250  1.3400e+07
; C  - OM       1000
;
#define gb_6        0.1320  1.2000e+07
; CR1  -  NR (6-ring)   1000
;
#define gb_7        0.1330  8.8700e+06
; H  -  S       750
;
#define gb_8        0.1330  1.0600e+07
; C  -  NT, NL  900
;
#define gb_9        0.1330  1.1800e+07
; C, CR1  -  N, NR, CR1, C (peptide, 5-ring)       1000
;
#define gb_10       0.1340  1.0500e+07
```

```
; C  -  N, NZ, NE       900
;
#define gb_11       0.1340  1.1700e+07
; C  -  NR (no H) (6-ring)      1000
;
#define gb_12       0.1360  1.0200e+07
; C  -  OA      900
;
#define gb_13       0.1380  1.1000e+07
; C  -  NR (heme)       1000
;
#define gb_14       0.1390  8.6600e+06
; CH2  -  C, CR1 (6-ring)       800
;
#define gb_15       0.1390  1.0800e+07
; C, CR1  -  CH2, C, CR1 (6-ring)       1000
;
#define gb_16       0.1400  8.5400e+06
; C, CR1, CH2  -  NR (6-ring)   800
;
#define gb_17       0.1430  8.1800e+06
; CHn  -  OA    800
;
#define gb_18       0.1430  9.2100e+06
; CHn  -  OM    900
;
#define gb_19       0.1435  6.1000e+06
; CHn  -  OA (sugar)    600
;
#define gb_20       0.1470  8.7100e+06
; CHn  -  N, NT, NL, NZ, NE     900
;
#define gb_21       0.1480  5.7300e+06
; CHn  -  NR (5-ring)   600
;
#define gb_22       0.1480  7.6400e+06
; CHn  -   NR (6-ring)  800
;
#define gb_23       0.1480  8.6000e+06
; O, OM  -   P     900
;
```

```
#define gb_24       0.1500  8.3700e+06
; O  -  S       900
;
#define gb_25       0.1520  5.4300e+06
; CHn  -   CHn (sugar)  600
;
#define gb_26       0.1530  7.1500e+06
; C, CHn  -   C, CHn    800
;
#define gb_27       0.1610  4.8400e+06
; OA  -   P     600
;
#define gb_28       0.1630  4.7200e+06
; OA  -   SI    600
;
#define gb_29       0.1780  5.9400e+06
; CH3  -   S    900
;
#define gb_30       0.1830  5.6200e+06
; CH2  -   S    900
;
#define gb_31       0.1870  3.5900e+06
; CH1  -   SI   600
;
#define gb_32       0.1980  0.6400e+06
; NR  -   FE    120
;
#define gb_33       0.2040  5.0300e+06
; S  -   S      1000
;
#define gb_34       0.2000  0.6280e+06
; NR (heme)  -   FE     120
;
#define gb_35       0.1000  2.3200e+07
; HWat  -   OWat        1110
;
#define gb_36       0.1100  1.2100e+07
; HChl  -   CChl        700
;
#define gb_37       0.1758  8.1200e+06
; CChl  -   CLChl       1200
```

```
;
#define gb_38       0.1530  8.0400e+06
; ODmso  -   SDmso      900
;
#define gb_39       0.1950  4.9500e+06
; SDmso  -   CDmso      900
;
#define gb_40       0.1760  8.1000e+06
; CCl4  -   CLCl4       1200
;
#define gb_41     0.163299  8.7100e+06
; HWat  -   HWat        1110
;
#define gb_42     0.233839  2.6800e+06
; HChl  -   CLChl        700
;
#define gb_43     0.290283  2.9800e+06
; CLChl -   CLChl       1200
;
#define gb_44     0.280412  2.3900e+06
; ODmso -   CDmso        900
;
#define gb_45     0.292993  2.1900e+06
; CDmso -   CDmso        900
;
#define gb_46     0.198842  3.9700e+06
; HMet  -   CMet         750
;
#define gb_47     0.287407  3.0400e+06
; CLCl4 -   CLCl4       1200
;
#define gb_48       0.1440  8.0400e+06
; GAGs - SULFATO - SDSMO e ODMSO
;
#define gb_49       0.1560  6.1000e+06
; GAGs - SULFATO - SDSMO e OA
;
#define gb_50       0.1630  2.2932e+06
; GAGs - SULFATO - SDSMO e N
;
#define gb_52       0.2540  0.6280e+06
```

```
; NR ()      -   CA-coor  120
;
;---
;       Table 2.5.3.1.
;       GROMOS bond-angle bending parameters
;
;  ICT(H)[N]  CT[N]  (T0[N])
;
#define ga_1         90.00      420.00
; NR(heme)  -  FE  -  NR(heme)  100
;
#define ga_2         96.00      405.00
; H  -  S  -  CH2       95
;
#define ga_3        100.00      475.00
; CH2  -  S  -  CH3     110
;
#define ga_4        103.00      420.00
; OA  -  P  -  OA       95
;
#define ga_5        104.00      490.00
; CH2  -  S  -  S       110
;
#define ga_6        108.00      465.00
; NR, C, CR1(5-ring)    100
;
#define ga_7        109.50      285.00
; CHn  - CHn - CHn, NR(6-ring) (sugar)  60
;
#define ga_8        109.50      320.00
; CHn, OA  - CHn  - OA, NR(ring) (sugar)        68
;
#define ga_9        109.50      380.00
; H -  NL, NT  -  H, CHn  - OA  - CHn(sugar)    80
;
#define ga_10       109.50      425.00
; H  -  NL  -  C, CHn          H  -  NT  -  CHn 90
;
#define ga_11       109.50      450.00
; X  -  OA, SI  -  X    95
;
```

```
#define ga_12       109.50      520.00
; CHn,C  -  CHn  -  C, CHn, OA, OM, N, NE       110
;
#define ga_13       109.60      450.00
; OM  -  P  -  OA       95
;
#define ga_14       111.00      530.00
; CHn  -  CHn  -  C, CHn, OA, NR, NT, NL        110
;
#define ga_15       113.00      545.00
; CHn  -  CH2  -  S     110
;
#define ga_16       115.00       50.00
; NR(heme)  -  FE  - NR 10
;
#define ga_17       115.00      460.00
; H  -  N  -  CHn       90
;
#define ga_18       115.00      610.00
; CHn, C  -  C  -  OA, N, NT, NL        120
;
#define ga_19       116.00      465.00
; H  -  NE  -  CH2      90
;
#define ga_20       116.00      620.00
; CH2  -  N  -  CH1     120
;
#define ga_21       117.00      635.00
; CH3 -  N  -  C, CHn  - C  - OM        120
;
#define ga_22       120.00      390.00
; H  -  NT, NZ, NE  -  C        70
;
#define ga_23       120.00      445.00
; H  -  NT, NZ  -  H    80
;
#define ga_24       120.00      505.00
; H - N - CH3, H, HC - 6-ring, H - NT - CHn     90
;
#define ga_25       120.00      530.00
; P, SI  -  OA  -  CHn, P       95
```

```
;
#define ga_26       120.00      560.00
; N, C, CR1 (6-ring, no H)      100
;
#define ga_27       120.00      670.00
; NZ  -  C  -  NZ, NE   120
;
#define ga_28       120.00      780.00
; OM  - P  -  OM        140
;
#define ga_29       121.00      685.00
; O  -  C  -  CHn, C          CH3  -  N  -  CHn 120
;
#define ga_30       122.00      700.00
; CH1, CH2  -  N  -  C  120
;
#define ga_31       123.00      415.00
; H  - N  - C   70
;
#define ga_32       124.00      730.00
; O  - C  - OA, N, NT, NL   C - NE - CH2        120
;
#define ga_33       125.00      375.00
; FE  - NR  - CR1 (5-ring)      60
;
#define ga_34       125.00      750.00
; -     120
;
#define ga_35       126.00      575.00
; H, HC  - 5-ring       90
;
#define ga_36       126.00      640.00
; X(noH)  - 5-ring      100
;
#define ga_37       126.00      770.00
; OM  - C  - OM 120
;
#define ga_38       132.00      760.00
; 5, 6 ring connnection 100
;
#define ga_39       155.00     2215.00
```

```
; SI  - OA  - SI        95
;
#define ga_40       109.50      434.00
; HWat  - OWat  - HWat  92
;
#define ga_41       107.57      484.00
; HChl  - CChl  - CLChl 105
;
#define ga_42       111.30      632.00
; CLChl  - CChl  - CLChl        131
;
#define ga_43        97.40      469.00
; CDmso  - SDmso  - CDmso       110
;
#define ga_44       106.75      503.00
; CDmso  - SDmso  -  ODmso      110
;
#define ga_45       108.53      443.00
; HMet  - OMet  - CMet  95
;
#define ga_46       109.50      618.00
; CLCl4  - CCl4  - CLCl4        131
;
#define ga_47       104.00      444.44
; angulo prodrg
;
;       Table 2.5.4.1
;       GROMOS improper (harmonic) dihedral angle parameters
;
; ICQ(H)[N] CQ[N] (Q0[N])
;
#define gi_1           0.0   167.42309
; planar groups 40
;
#define gi_2      35.26439   334.84617
; tetrahedral centres   80
;
#define gi_3           0.0   669.69235
; heme iron     160
;
#define gi_4          -2.2   334.84617
```

```
; DIH1_1C4      160
;
#define gi_5         -27.0   334.84617
; DIH2_1C4      160
;
#define gi_6          1.4    334.84617
; DIH3_1C4      160
;
#define gi_7          2.2    334.84617
; DIH1_4C1      160
;
#define gi_8         27.0    334.84617
; DIH2_4C1      160
;
#define gi_9         -1.4    334.84617
; DIH3_4C1      160
;
#define gi_10         0.0    209.30000
; diedro improprio - prodrg
;
#define gi_11        24.5    334.84617
; DIH1_2SO      160
;
#define gi_12        43.0    334.84617
; DIH2_2SO      160
;
#define gi_13        31.0    334.84617
; DIH3_2SO      160
;
#define gi_14       -10.3    334.84617
; DIH1_1T2-furanose
;
#define gi_15       -42.0    334.84617
; DIH2_1T2-furanose
;
#define gi_16        11.6    334.84617
; DIH3_1T2-furanose
;
#define gi_17        54.3    334.84617
; DIH4_1T2-furanose
;
```

```
#define gi_18         30.0   334.84617
; DIH1_4T3-furanose
;
#define gi_19         37.8   334.84617
; DIH2_4T3-furanose
;
#define gi_20         25.7   334.84617
; DIH3_4T3-furanose
;
#define gi_21        -12.0   334.84617
; DIH4_4T3-furanose
;
;       Table 2.5.5.1
;       GROMOS (trigonometric) dihedral torsional angle parameters
;
; ICP(H)[N]  CP[N] PD[N] NP[N]
;
#define gd_1    180.000       5.86          2
; -C-C- 1.4
;
#define gd_2    180.000       7.11          2
; -C-OA- (at ring)      1.7
;
#define gd_3    180.000       16.7          2
; -C-OA- (carboxyl)     4.0
;
#define gd_4    180.000       33.5          2
; -C-N, NT, NE, NZ,NR-  8.0
;
#define gd_5    180.000       41.8          2
; -C-CR1- (6-ring)      10.0
;
#define gd_6      0.000        0.0          2
; -CH1 (sugar)-NR(base)-        0.0
;
#define gd_7      0.000      0.418          2
; O-CH1-CHn-no O        0.1
;
#define gd_8      0.000       2.09          2
; O-CH1-CHn-O   0.5
;
```

```
#define gd_9      0.000       3.14          2
; -OA-P-         0.75
;
#define gd_10     0.000       16.7          2
; -S-S- 4.0
;
#define gd_11     0.000       1.05          3
; -OA-P-         0.25
;
#define gd_12     0.000       1.26          3
; -CHn-OA(no sugar)-    0.3
;
#define gd_13     0.000       2.93          3
; -CH2-S-        0.7
;
#define gd_14     0.000       3.77          3
; -C,CHn,SI-NT,NL,OA(sugar)-    0.9
;
#define gd_15     0.000       4.18          3
; HC-C-S-        1.0
;
#define gd_16     0.000       5.44          3
; HC-C-C-        1.3
;
#define gd_17     0.000       5.86          3
; -CHn,SI-CHn-  1.4
;
#define gd_18     0.000        0.0          4
; -NR-FE-        0.0
;
#define gd_19   180.000        1.0          6
; -CHn-N,NE-     0.24
;
#define gd_20     0.000        1.0          6
; -CHn-C,NR (ring), CR1-        0.24
;
#define gd_21     0.000       3.77          6
; -CHn-NT-      0.9
;

; get the constraint distances for dummy atom constructions
```

```
#include "ff_dum.itp"

[ constrainttypes ]
; now the constraints for the rigid NH3 groups
 MNH3    C    2   DC_MNC1
 MNH3  CH1    2   DC_MNC2
 MNH3  CH2    2   DC_MNC2
 MNH3 MNH3    2   DC_MNMN
; and the angle-constraints for OH and SH groups in proteins:
  CH2    H    2   DC_CO
  CH1    H    2   DC_CO
    C    H    2   DC_CO
    P    H    2   DC_PO

; bond-, angle- and dihedraltypes for specbonds:
[ bondtypes ]
S      S       2    gb_33
NR     FE      2    gb_32
; cystine - heme link (is CR1-S, use CH2-S):
S      CR1     2    gb_30

[ angletypes ]
CH1    CH2    S     2   ga_15
CH2    S      S     2   ga_5
CR1    NR    FE     2    ga_33
NR     FE    NR     2   ga_16
; cystine - heme link (is CH2-S-CR1, use CH2-S-CH2):
CH2    S     CR1    2   ga_3
; cystine - heme link (is S-CR1-C/CH2, use CHn-CH2-S):
S      CR1   C      2   ga_15
S      CR1   CH2    2   ga_15

[ dihedraltypes ]
S      S      1   gd_10
NR     FE     1   gd_18
CH2    S      1   gd_13
; cystine - heme link (is CH2-S-CR1-C, use -CH2-S-):
CR1    S      1   gd_13
```

9 Curriculum Vitae

I. Formação Acadêmica/Titulação

2007 Doutorado em Biologia Celular e Molecular.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Descrição estrutural e conformacional da poliproteína gag e de glicoproteínas do HIV-1: relação dos componentes virais à maturação do virion e ao processo infeccioso
Orientador: Hugo Verli
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

2004 - 2006 Mestrado em Ciências Biológicas (Bioquímica).
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Modelagem molecular e estudos de Docking da enzima Corismato Sintase de Mycobacterium tuberculosis, Ano de obtenção: 2006
Orientador: Luiz Augusto Basso
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

2002 - 2004 Especialização em Farmácia Bioquímica.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Perfil glicêmico da Diabete tipo MODY-2 na gestação: descrição de um caso
Orientador: Carmen Regla Vargas

1996 - 2001 Graduação em Farmácia.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Modelagem molecular da EPSP sintase de Mycobacterium tuberculosis por modelagem comparativa por homologia
Orientador: Diógenes Santiago Santos

II. Formação complementar

2005 - 2005 Curso de curta duração em Treinamento no Banco de Dados Hiv Base.
Fiocruz Centro de Pesquisas Gonçalo Monuz, FIOCRUZ/CPQGM, Brasil

2002 - 2002 Extensão universitária em Linux Básico.
Pontifícia Universidade Católica do Rio Grande do Sul, PUCRS, Porto Alegre, Brasil

2002 - 2002 Extensão universitária em Escola de modelagem molecular.
Laboratório Nacional de Computação Científica, LNCC/MCT, Brasil

2001 - 2001 Curso de curta duração em Introdução Ao Fortran 90.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil

III. Artigos completos publicados em periódicos

1. 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.

2. THOMPSON, CLAUDIA E., FERNANDES, CLÁUDIA L., NORBERTO DE SOUZA, OSMAR, FREITAS, LORETA B., SALZANO, FRANCISCO M.
Evaluation of the impact of functional diversification on Poaceae, Brassicaceae, Fabaceae, and Pinaceae alcohol dehydrogenase enzymes. Journal of Molecular Modeling. , p.1 - , 2009.

3. POL-FACHIN, L., FERNANDES, C. L., VERLI, H.
GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles

through molecular dynamics simulations. Carbohydrate Research. , v.344, p.491 - 500, 2009.

4. FERNANDES, C. L., BREDA, Ardala, SANTOS, D. S., BASSO, L. A., SOUZA, O. N.
A structural model for chorismate synthase from Mycobacterium tuberculosis in complex with coenzyme and substrate. Computers in Biology and Medicine. , v.37, p.149 - 158, 2007.

5. FONSECA, I. O., SILVA, R. G., FERNANDES, C. L., SOUZA, O. N., BASSO, L. A., SANTOS, D. S.
Kinetic and chemical mechanisms of shikimate dehydrogenase from Mycobacterium. Archives of Biochemistry and Biophysics. , v.457, p.123 - 133, 2007.

6. THOMPSON, CLAUDIA E., FERNANDES, C. L., DE SOUZA, OSMAR N., SALZANO, FRANCISCO M., BONATTO, SANDRO L., FREITAS, LORETA B.
Molecular Modeling of Pathogenesis-Related Proteins of Family 5. Cell Biochemistry and Biophysics. , v.44, p.385 - 394, 2006.

7. BREDA, ARDALA, FERNANDES, C. L., ALMEIDA, SABRINA ESTEVES DE MATOS, FRANCO, H. M., ROSSETTI, MARIA LÚCIA ROSA, RODRIGUES, ROSÂNGELA, BRÍGIDO, LUÍS FERNANDO, CORTEZ-HERRERA, ELIZABETH
Bioinformatics Tools for HIV-1 Identification in Southern Brazilian States. Lecture Notes in Computer Science. , v.3594, p.234 - 237, 2005.

8. FERNANDES, C. L., SANTOS, D. S., BASSO, L. A., SOUZA, O. N.
Structure Prediction and Docking Studies of Chorismate Synthase from Mycobacterium. Lecture Notes in Computer Science. , v.3594, p.118 - 127, 2005.

IV. **Artigos aceitos para publicação**

1. LAZZAROTTO, F., TEIXEIRA, F. K., ROSA, S. B., DUNAND, C., FERNANDES, C.L., FONTENELE, A. V., SILVEIRA, J. A. G., VERLI, H., MARGIS, R., MARGIS-PINHEIRO, M.
APX-R is a new heme-containing protein functionally associated to APx but evolutionarily divergent.. New Phytologist (Print). , 2011.

V. **Trabalhos publicados em anais de eventos (resumo)**

1. FERNANDES, C. L., VERLI, H.
Effects of glycosilation in the human alpha1-acid glycoprotein (AGP) structure In: V Escola de Modelagem Molecular em Sistemas Biológicos, 2010, Petrópolis-RJ.
V Escola de Modelagem Molecular em Sistemas Biológicos. , 2010.

2. FERNANDES, C. L., VERLI, H.
Effects of glycosylation in the human alpha1-Acid glycoprotein (AGP) conformation In: XXXVIII annual Meeting of SBBq, 2009, Águas de Lindóia-SP.
XXXVIII annual Meeting. , 2009.

3. CHIODI, C. G., FERNANDES, C. L., VERLI, H.
Structural and conformational characterization of human prothrombin In: XXXVIII annual Meeting of SBBq, 2009, Águas de Lindóia-SP.
XXXVIII annual Meeting. , 2009.

4. FERNANDES, C. L., POL-FACHIN, L., BERGAMINI, G., VERLI, H.
Prediction of Disaccharides Conformational Ensembles in Solution from Sequence In: XXXVII Reunião Anual da SBBq, 2008, Águas de Lindóia/SP.
XXXVII Reunião Anual da SBBq. , 2008.

5. POL-FACHIN, L., FERNANDES, C. L., VERLI, H.
Prediction of glycoproteins conformational ensembles in solution In: XXXVII Reunião Anual da SBBq, 2008, Águas de Lindóia-SP.
XXXVII Reunião Anual da SBBq. , 2008.

6. FERNANDES, C. L., BECKER, C. F., GIESEL, G. M., LIMA, M., VERLI, H.
The use of molecular dynamics simulations in the refinement of crystallografic structures In: XXXVI Annual meeting of the brazilian society for biochemistry and molecular biology, 2007, Salvador.
XXXVI Annual Meeting Program 2007. , 2007. p.106 – 106.

7. ALMEIDA, SABRINA ESTEVES DE MATOS, SPERHACKE, R. D., BREDA, ARDALA, FERNANDES, C. L., ROSSETTI, MARIA LÚCIA ROSA, CORTEZ-HERRERA, ELIZABETH, BACIN, T., RODRIGUES, ROSÂNGELA, BRÍGIDO, LUÍS FERNANDO
Hiv-1 subtype analysis at Rio Grande do Sul In: XXIII Congresso Brasileiro de Microbiologia, 2005, Santos- SP.

8. THOMPSON, C. E., FERNANDES, C. L., SOUZA, O. N., SALZANO, Francisco Mauro, MENEGHINI, Rogério, FREITAS, Loreta Brandão de
Comparative Modelling by Homology in the study of the Molecular Evolution of the Pathognesis-Related Protein 5 In: First Latin American Protein Society Meeting, 2004, Angra dos Reis.

9. THOMPSON, C. E., FERNANDES, C. L., SOUZA, O. N., SALZANO, Francisco Mauro, BONATTO, Sandro Luís, FREITAS, Loreta Brandão de
Métodos de Máxima Verossimilhança e Modelagem Molecular Comparativa por Homologia no estudo da Evolução Molecular das Proteínas de Defesa do Grupo PR5 em Plantas. In: XIV Encontro de Geneticistas do Rio Grande do Sul, 2004, Porto Alegre.

10. FERNANDES, C. L., SANTOS, D. S., BASSO, L. A., SOUZA, O. N.
STRUCTURE PREDICTION AND DOCKING STUDIES OF CHORISMATE SYNTHASE FROM Mycobacterium tuberculosis In: SBBq 2004 - XXXIII Reunião Anual, 2004, Caxambu -MG.
Programa e Resumos. , 2004.

11. FERNANDES, C. L., SCHROEDER, E. K., SANTOS, D. S., BASSO, L. A., SOUZA, O. N.
HOMOLOGY MODELING OF KasA AND KasB ENZYMES FROM Mycobacterium tuberculosis AND INHIBITOR DOCKING STUDIES WITH CERULENIN AND THIOLACTOMYCIN In: V Ibero-American Congress of Biophysics, 2003, Rio de Janeiro -RJ.
V Ibero-American Congress of Biophysics. , 2003.

12. FERNANDES, C. L., SCHROEDER, E. K., SANTOS, D. S., BASSO, L. A., SOUZA, O. N.
STRUCTURE PREDICTION AND INHIBITOR DOCKING STUDIES OF KasA AND KasB ENZYMES FROM Mycobacterium tuberculosis In: SBBq 2003 - XXXII Reunião Anual, 2003, Caxambú - MG.
Programa e Resumos. , 2003.

13. KOLBERG, M. L., FERNANDES, C. L., SCHROEDER, E. K., BASSO, L. A., SANTOS, D. S., SOUZA, O. N.
EXPLORANDO GENÔMICA E BIOINFORMÁTICA PARA DESCOBRIR DOMÍNIOS FUNCIONAIS NA CORISMATO SINTASE DE Mycobacterium tuberculosis In: XII Encontro de geneticistas do Rio Grande do Sul, 2002, Porto Alegre.

14. FERNANDES, C. L., OLIVEIRA, J. S., SCHROEDER, E. K., BASSO, L. A., SANTOS, D. S., SOUZA, O. N.
Genômica e Bioinformática estrutural: utilizando a EPSP Sintase de Mycobacterium tuberculosis como alvo para o desenho de novos fármacos antituberculose In: XII Encontro de geneticistas do Rio Grande do Sul, 2002, Porto Alegre.

15. FERNANDES, C. L., OLIVEIRA, J. S., SCHROEDER, E. K., BASSO, L. A., SANTOS, D. S., SOUZA, O. N.
Genômica e bioinformática estrutural: utilizando a EPSP Sintase de Mycobacterium tuberculosis como alvo para o desenho de novos fármacos antituberculose In: Escola de modelagem molecular em sistemas biológicos, 2002, Rio de Janeiro.

VI. **Participação em banca de trabalhos de conclusão**

Graduação

1. VERLI, H., FERNANDES, C. L.
Participação em banca de Eduarda Schultze. **Análise conformacional das propriedades anticoagulantes de saponinas**, 2008
(Biomedicina) Universidade Federal do Rio Grande do Sul

2. VERLI, H., FERNANDES, C. L.
Participação em banca de Laércio Pol Facchin. **Characterization of the structure and flexibility of heparin using molecular dynamics simulations**, 2007
(Biomedicina) Universidade Federal do Rio Grande do Sul

VII. Orientações de Iniciação Científica:

a. Concluídas:

Carla Gottschald Chiodi. Centro de Biotecnologia, UFRGS, a partir de julho de 2008 até dezembro de 2010.

b. Em andamento:

Pablo Ricardo Arantes. Centro de Biotecnologia, UFRGS, a partir de abril de 2010.

VIII. Bolsa recebida:

Bolsista CAPES do programa de doutorado em biologia molecular e celular pelo Centro de Biotecnologia da UFRGS a partir de março de 2007.