UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

LEANDRO OLIVEIRA BORTOT

Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores

Ribeirão Preto
2018

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas para obtenção do Título de Doutor em Ciências

Área de Concentração: Química e Física Biológica.

Orientado: Leandro Oliveira Bortot

Orientador: Prof. Dr. Antonio Caliri

Versão corrigida da Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas em 16/05/2018. A versão original encontra-se disponível na Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP.

Ribeirão Preto
2018

BORTOT, L.O.

Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores

DOUTORADO
FCFRP/USP
2018

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Bortot, Leandro Oliveira
Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores. Ribeirão Preto, 2018.
127 p. : il. ; 30cm.

Tese de Doutorado, apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP – Área de concentração: Química e Física Biológica.
Orientador: Caliri, Antônio

1. Dengue. 2. Dinâmica molecular. 3. Energia livre.

FOLHA DE APROVAÇÃO

Leandro Oliveira Bortot
Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas da Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP para obtenção do Título de Doutor em Ciências

Área de Concentração: Química e Física Biológica.

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura:_____

Prof. Dr. _____

Instituição: _____ Assinatura:_____

Prof. Dr. _____

Instituição: _____ Assinatura:_____

Prof. Dr. _____

Instituição: _____ Assinatura:_____

Prof. Dr. _____

Instituição: _____ Assinatura:_____

Dedico esta tese aos meus pais, que me deram todo o apoio necessário, aos professores que me incentivaram e aos amigos que acompanharam.

Agradecimentos

Ao Professor Dr. Antônio Caliri por todas as discussões e pelo companheirismo ao longo dos anos, bem como aos ex colegas de laboratório por todas as experiências vividas: Renata Almeida Garcia Reis, Flávio Henrique Sant'Ana Costa, João Paulo dal Molin, Ricardo Oliveira dos Santos Soares e Rodrigo Antônio Faccioli.

À Universidade de São Paulo, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Programa de Pós-Graduação em Ciências Farmacêuticas pelo suporte ao longo do doutorado.

À FAPESP pelo apoio financeiro através do projeto regular de doutorado (processo 2013/00927-7) e pela bolsa BEPE (processo 2016/04958-2), possibilitando minha dedicação intensa à pesquisa ao longo do doutorado.

Ao Professor David van der Spoel por ter me recebido gentilmente em seu grupo de pesquisa na Universidade de Uppsala, Suécia, e aos antigos colegas de grupo durante pouco menos de 1 ano pelo ambiente amigável e estimulante: Zahedeh Bashardanesh, Mohammad Ghahremanpour, Benjamin Schroeder, Nina Fischer, Marie-Madeleine Walz, Samuel Ferrer e Sergio Manzetti.

À USP, Rice University, HPC2N e PDC pelo acesso aos vários recursos computacionais usados ao longo deste trabalho.

Aos meus pais por apoiar a minha carreira na pesquisa.

Aos amigos que me acompanham.

“We choose to go to the Moon! (…) and do the other things,
not because they are easy, but because they are hard;
because that goal will serve to organize and measure the best of our energies and skills,
because that challenge is one that we are willing to accept,
one we are unwilling to postpone, and one we intend to win”

–John F. Kennedy, 12 de Setembro de 1962

RESUMO

BORTOT, L.O. **Mecanismos moleculares de reconhecimento das glicosilações do envelope do vírus da Dengue pelas lectinas tipo-C e seus potenciais inibidores**. 2018. Tese (Doutorado). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

Dengue é uma doença tropical negligenciada que atualmente ameaça mais da metade de população mundial e representa custo anual de bilhões de dólares para as áreas afetadas. Avanços recentes na elucidação estrutural de proteínas humanas e virais contribuíram para o entendimento do ciclo de replicação viral e das interações vírus–hospedeiro a nível molecular. Em particular, a camada mais externa do vírus é composta por 180 monômeros da glicoproteína do envelope. Cada um desses monômeros apresenta duas glicosilações que são reconhecidas por lectinas cuja atividade depende de Ca^{2+}. A interação entre o vírus e essas lectinas favorece a infecção ou o aparecimento de sintomas mais severos. Neste trabalho aplicamos simulações de dinâmica molecular e métodos de estimativa de afinidade para avançar nosso conhecimento sobre os mecanismos moleculares do reconhecimento de glicosilações *high-mannose* pelas lectinas DC-SIGN e MR, ambas já experimentalmente validadas como alvos biológicos para desenvolvimento de novos antivirais. Adicionalmente, através de *virtual screening* usando um conjumto de programas de implementação própria e uma biblioteca de moléculas já aprovadas para uso como fármacos, encontramos uma molécula (Iohexol) que apresenta alto potencial de interação com ambos os receptores. Embora testes experimentais ainda sejam necessários para validar esse achado, nossos resultados sugerem que essa molécula, e eventualmente moléculas similares, podem agir como inibidor da infecção do vírus da Dengue por um mecanismo duplo.

Palavras-chave: Dengue, Dinâmica molecular, Energia livre

ABSTRACT

BORTOT, L.O. **Molecular mechanisms underlying the recognition of the Dengue virus envelope glycosylations by C-Type Lectins and its potential inhibitors.** 2018. Thesis (Doctoral). Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

Dengue is a tropical neglected disease that currently threatens more than half the world's population and represents a yearly cost of billions of dollars to the affected areas. Recent advances in the elucidation of 3D structures of human and viral proteins has contributed to the understanding of the viral replication cycle and virus–host interactions at the molecular level. In particular, the outermost layer of the virus is composed by 180 monomers of the envelope glycoprotein. Each one of these monomers displays two glycosylations that are recognized by lectins which have Ca^{2+}-dependent activity. The interaction between the virus and these lectins favors infection or severe disease onset. In this work we apply molecular dynamics simulations and affinity estimation methods to advance our knowledge about the molecular mechanisms underlying the recognition of the high-mannose glycosylation by the DC-SIGN and MR lectins, both of which are already validated as targets for the development of new antivirals against Dengue. Additionally, by running virtual screening assays using a set of programs implemented by us and a library containing molecules which are already approved to be used as drugs, we found one molecule (Iohexol) which presents high potential of interaction with both receptors. Although experimental testing is still necessary to validate this finding, our results suggest that this molecule, and eventually similar ones, can act as an inhibitor of the Dengue virus infection by a dual mechanism.

Keywords: Dengue, Molecular dynamics, Free energy

LISTA DE FIGURAS

Figura 1: Representação esquemática do ciclo de replicação do vírus da Dengue. A etapa de adesão e entrada está destacada dentro da elipse vermelha. Figura adaptada da referência 15 4

Figura 2: Dímero da glicoproteína do envelope viral. A proteína está mostrada em cartoon. Os domínios I, II e III estão mostrados em vermelho, amarelo e azul, respectivamente. As glicosilações *high-mannose* estão mostradas como *sticks* verdes. 5

Figura 3: Rotâmeros da ligação do tipo α1,6. Figura extraída da referência 33. 6

Figura 4: Representação esquemática do processo de *trimming* e processamento da glicosilação *high-mannose* inicial, no retículo endoplasmático, até a *complex-type* no trans-Golgi. 8

Figura 5: **a)** Representação esquemática da estrutura completa do DC-SIGN. Os CRDs estão representados em azul e os domínios do pescoço em verde. Figura adaptada da referência 68. **b)** Estrutura 3D à baixa resolução da porção extracelular do DC-SIGN. Figura adaptada da referência 72. 11

Figura 6: Estruturas cristalográficas do DC-SIGN. A proteína está mostrada em cartoons verdes, os Ca^{2+} como esferas azuis e os carboidratos como *sticks* amarelos. **a)** O sítio de complexação principal de Ca^{2+} está destacado pela seta preta. O sítio auxiliar, que complexa dois Ca^{2+}, está destacado dentro do círculo preto; **b)** Estrutura do complexo com o fragmento Man_4 (PDB ID 1SL4) **c)** Estrutura do complexo com o fragmento Le^x (PDB ID 1SL5). 12

Figura 7: **a)** Representação esquemática da estrutura completa do MR. Adaptado da referência 79. **b)** Estrutura cristalográfica do CTLD-4 do MR no qual aconteceu troca de domínios entre unidades assimétricas vizinhas no cristal, que estão representados em cores diferentes. Figura adaptada da referência 90. 14

Figura 8: **a)** Representação esquemática da estrutura completa do CLEC5A. Seu CLTD está mostrado em vermelho, o domínio linker como uma linha e sua hélice transmembrana está em roxo. A interação com DAP12 ocorre dentro da membrana. Figura adaptada da referência 100. **b)** Estrutura cristalográfica do seu CTLD (PDB ID 2YHF). 16

Figura 9: Ciclo termodinâmico esquemático para o método MM/PBSA. O solvente está representado pelo fundo azul e o vácuo pelo fundo preto. Figura adaptada de "*AMBER Advanced Tutorials – Tutorial 3*". 22

Figura 10: Força que a mola virtual impõe sobre o sistema receptor–ligante para separá-los durante uma trajetória representativa de simulação de dinâmica molecular dirigida. A força aumenta ao longo to tempo e alcança um máximo, que é a força de ruptura, antes de romper as interações receptor–ligante. 23

Figura 11: Representação esquemática do método *Umbrella Sampling*. **Linha superior:** Amostragem inicial da coordenada de reação através de simulações dirigidas. **Linha central:** Simulações de amostragem individuais nas quais são aplicados potenciais harmônicos para restringir a amostragem do caminho de reação dentro de cada janela. **Linha inferior:** Histogramas de amostragem com boa sobreposição. Fonte: "*GROMACS Tutorial – Umbrella Sampling*". 24

Figura 12: Estruturas de referência dos três receptores considerados: **a)** DC-SIGN; **b)** MR; **c)** CLEC5A. A proteína está mostrada em cartoon verde e os Ca2+ como esferas azuis. O Ca2+ principal está destacado por uma seta preta. O sítio de complexação auxiliar de Ca2+ do DC-SIGN está destacado

dentro de um círculo preto. Os dois Ca2+ presentes no sítio auxiliar do DC-SIGN podem ainda ser individualmente identificados como “auxiliar interno” e “auxiliar externo”, destacados por uma seta cinza e branca, respectivamente. 30

Figura 13: Interação de Ca^{2+} da solução durante simulações nas quais os Ca^{2+} cristalográficos foram removidos e $CaCl_2$ foi usado para neutralizar as cargas e ajustar a força iônica em 150 mM. (a,d) DC-SIGN; (b,e) MR; e (c,f) CLEC5A. Linha superior: estruturas cristalográficas, incluindo os Ca^{2+}, representados por esferas azuis. Linha inferior: ocupância da interação de cada resíduo com Ca^{2+} da solução projetada sobre a estrutura cristalográfica. A escala de cores vai de branco representando 0% de ocupância —ausência total de interação entre o resíduo e Ca^{2+}— até vermelho representando 100% de ocupância —interação do resíduo com Ca^{2+} durante todo o tempo de simulação analisado. 32

Figura 14: Energia potencial receptor–cátion dos complexos envolvendo os receptores DC-SIGN, MR e CLEC5A com diferentes quantidades de Ca^{2+} complexados a eles, conforme indicado no topo do gráfico: 0: nenhum Ca^{2+} complexado ao CTLD; 1: um Ca^{2+} complexado ao sítio principal; 2: dois Ca^{2+} ambos complexados ao sítio auxiliar; 3: três Ca^{2+} complexados ao CTLD, sendo um ao sítio principal e dois ao sítio auxiliar simultaneamente. Enquanto o tamanho total das barras representa a energia potencial do complexo receptor–cátions, suas subdivisões representam as contribuições da energia potencial intraproteína (área preta), interações proteína–Ca^{2+} (área cinza) e interações proteína–Na^{+} (área hachurada). 34

Figura 15: Flexibilidade total dos receptores conforme descrita pela somatória do RMSF dos carbonos-α de todos os seus resíduos. O estado de ocupação dos sítios de complexação de Ca^{2+} como descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas). 35

Figura 16: Diferença de flexibilidade de cada resíduo, $\Delta RMSF_{resíduo}$, após complexação de Ca^{2+}. Valores negativos indicam que o resíduo é menos flexível no estado saturado de Ca^{2+} do que no estado livre de Ca^{2+}. Áreas coloridas correspondem a segmentos de estrutura secundária com α–hélices indicadas em ciano e folhas–β em verde. **Painel superior:** Diferença de flexibilidade entre DC-SIGN complexando três e zero Ca^{2+}. **Painel inferior:** Diferença de flexibilidade entre MR complexando um e zero Ca^{2+}. As altas incertezas, indicadas pelas barras de erro, são consequência de variações estruturais no estado livre de Ca^{2+}. 36

Figura 17: Integridade estrutural do sítio principal do DC-SIGN e MR conforme descrita pela **(a)** somatória dos RMSFs dos átomos pesados dos resíduos que os compõem; e **(b)** RMSD dos átomos pesados dos resíduos que os compõem em relação à estrutura cristalográfica de referência. O estado de ocupação dos sítios de complexação de Ca^{2+} está descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas). 37

Figura 18: Integridade estrutural do sítio principal do **(a,b)** DC-SIGN e **(c,d)** MR conforme descrita pela **(a,c)** RMSF dos átomos pesados dos resíduos que os compõem; e **(b,d)** RMSD dos átomos pesados dos resíduos que os compõem em relação à estrutura cristalográfica de referência. O estado de ocupação dos sítios de complexação de Ca^{2+} conforme descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal

ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas). 38

Figura 19: Visualização da integridade estrutural do sítio principal do **(a,b)** DC-SIGN e **(c,d)** MR no estado saturado de Ca^{2+} e **(a,c)** na ausência de Ca^{2+}. *Snapshots* obtidos após mais de 150 ns de simulação de dinâmica molecular. A proteína está mostrada em *cartoon* verde com as cadeias laterais dos resíduos do sítio principal destacadas como *sticks* e as demais mostradas como linha. Ca^{2+} é mostrado como esferas azuis e Na^{+} como esperas cinza escuro. 39

Figura 20: Perfil de flexibilidade por resíduo, $RMSF_{resíduo}$. Áreas coloridas correspondem a segmentos de estrutura secundária com α–hélices indicadas em ciano e folhas–β em verde. **Painel superior:** DC-SIGN quando complexado a três Ca^{2+}. **Painel inferior:** CLEC5A. 40

Figura 21: **a)** Força de ruptura para remoção de cada Ca^{2+} do DC-SIGN, conforme especificado no eixo x. O estado de ocupação dos sítios de complexação de Ca^{2+} está descrito na legenda: apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios principal e auxiliar ocupados (barras pretas). **b)** Estrutura de referência do DC-SIGN. O sítio principal está indicado por uma seta preta e o auxiliar por um círculo preto. Os Ca^{2+} que fazem parte do sítio auxiliar são individualmente chamados de auxiliar interno e externo, indicados por uma seta cinza e branca, respectivamente. 41

Figura 22: **(a,c,e)** Contribuições para $\Delta G(\text{cátion})$ e $\Delta\Delta G(\text{cátion})$: mecânica molecualar (azul), solvatação polar (vermelho) e total (preto). O valor usado para a constante dielétrica interna da proteína foi 1. **(b,d,f)** $\Delta\Delta G(\text{cátion})$ em função da constante dielétrica interna do receptor. 44

Figura 23: Contribuições dos resíduos que compõem o sítio principal para a energia livre de interação CTL–carboidrato quando há Ca^{2+} (barras cinzas) ou Na^{+} no sítio principal (barras hachuradas) e a diferença entre tais contribuições (barras pretas). **(a)** DC-SIGN – Man_4; **(b)** DC-SIGN – Le^X; **(c)** MR – Man_4. 46

Figura 24: Densidade local de água ao redor do **(a, b, c)** DC-SIGN e **(d, e, f)** MR. **Primeira coluna:** sítios onde a densidade local de água é pelo menos 10 desvios padrão maior do que a média da caixa estão representados por pequenas esferas vermelhas. **Segunda coluna:** Sítios de alta densidade local de água próximos do Ca^{2+} principal, evidenciando a identificação da estrutura da sua primeira camada de solvatação. **Terceira coluna:** Sobreposição da estrutura cristalográfica do DC-SIGN em complexo com Man_4 à análise de densidade de água. As hidroxilas do anel do carboidrato que interage com o Ca^{2+} principal ocupam dois dos seus sítios de solvatação. 48

Figura 25: Estrutura da glicosilação *high-mannose*. Esquerda: Representação esquemática com os tipos de ligação glicosídica indicados; direita: representação atomística na mesma orientação usada na representação esquemática. Os resíduos de GlcNAc estão em cinza e cada manose está representada em uma cor diferente. A esfera amarela representa o ponto onde a glicosilação é covalentemente ligada à glicoproteína do envelope do vírus da Dengue. **a)** Glicosilação maior, $GlcNAc_2Man_9$. **b)** Glicosilação menor, $GlcNAc_2Man_5$. 49

Figura 26: Interação entre o fragmento Man_5, representado como *sticks* e superfície amarela, com o resíduo Phe-313 do DC-SIGN, destacado em *sticks* e esferas verdes. A proteína está mostrada em *cartoon* cinza e o Ca^{2+} principal como uma esfera azul. 50

Figura 27: Estrutura esquemática da glicosilação $GlcNAc_2Man_9$ e a numeração de cada resíduo de manose que adotamos. $GlcNAc_2Man_5$ não tem as manoses 3, 4, 7 e 9, destacadas em ciano. 51

Figura 28: Detalhes das ligações de hidrogênio formadas pelos resíduos **a)** Man-6 e **b)** Man-8 com DC-SIGN. A proteína está mostrada em *cartoon* verde com cadeias laterais em *sticks*, o fragmento de glicosilação em *sticks* amarelos e o Ca^{2+} principal como uma esfera azul. 51

Figura 29: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com os fragmentos da glicosilação *high-mannose* **(a,b)** Man_5 e **(c,d)** Man_9 mostrada em escala de cor gradual de branco, que corresponde a ocupância 0 %, até vermelho, que corresponde a ocupância 100%. A esfera azul é o Ca^{2+} principal. A coluna da direita mostra o mesmo da esquerda com a adição do fragmento de glicosilação **(b)** Man_5 ou **(d)** Man_9, mostrado como *sticks* amarelos. 54

Figura 30: **a)** Interação entre o fragmento Man_5, representado como *sticks* amarelos, com os resíduos His-692, Leu-694, Tyr-729, Ile-749 e His-753 do MR, destacados como *sticks* e esferas verdes. A proteína está mostrada em *cartoon* cinza e o Ca^{2+} principal como uma esfera azul. **b)** Detalhe do contato entre os anéis da manose principal do Man_5, em amarelo, e Tyr-729 do MR, em verde. 55

Figura 31: Estrutura esquemática da glicosilação $GlcNAc_2Man_9$ e a numeração de cada resíduo de manose que adotamos. $GlcNAc_2Man_5$ não tem as manoses 3, 4, 7 e 9, destacadas em ciano. 56

Figura 32: Detalhes das ligações de hidrogênio formadas pelos resíduos **a)** Man-5, **b)** Man-6, **c)** Man-8 e **d)** Man-9 das glicosilações *high-mannose*. 56

Figura 33: Ocupância das ligações de hidrogênio de cada resíduo do MR com os fragmentos da glicosilação *high-mannose* **(a,b)** Man_5 e **(c,d)** Man_9 mostrada em escala de cor gradual de branco, que corresponde a ocupância 0 %, até vermelho, que corresponde a ocupância 100%. A esfera azul é o Ca^{2+} principal. A coluna da direita mostra o mesmo da esquerda com a adição do fragmento de glicosilação **(b)** Man_5 ou **(d)** Man_9, mostrado como *sticks* amarelos. 59

Figura 34: PMFs calculados considerando vários intervalos das simulações de *Umbrella Sampling* para a interação entre o CRD do DC-SIGN e **a)** $GlcNAc_2Man_5$; **b)** $GlcNAc_2Man_9$; **c)** Le^x. Para cada complexo são mostrados dois gráficos: à esquerda são mostrados PMFs calculados considerando intervalos de 10 ns que não se sobrepõem. À direita são mostrados PMFs calculados a partir de um determinado tempo que varia de 0 até 40 ns em passos de 10 ns até o fim da simulação, 50 ns. 61

Figura 35: **a)** PMF calculado pelo método *Umbrella Sampling* para descrever a interação entre o CRD do DC-SIGN e as glicosilações $GlcNAc_2Man_5$, $GlcNAc_2Man_9$ e Le^x. Histogramas de amostragem para as glicosilações: **b)** $GlcNAc_2Man_5$, **c)** $GlcNAc_2Man_9$ e **d)** Le^x. 62

Figura 36: Ocupância das ligações de hidrogênio entre DC-SIGN e as glicosilações *high-mannose* projetada sobre sua estrutura 3D. A escala de cores vai de branco a vermelho para indicar 0 e 100% de ocupância, respectivamente. A esfera azul é o Ca^{2+} principal. **Linha superior:** $GlcNAc_2Man_5$, sendo **a)** conformação construída a partir da estrutura cristalográfica de referência; **b)** rotâmero gt-gg interagindo com o Ca^{2+} principal pela Man-1; **c)** rotâmero gt-gt interagindo com o Ca^{2+} principal pela Man-3; **d)** rotâmero gg-gt interagindo com o Ca^{2+} principal pela Man-7. **Linha central:** $GlcNAc_2Man_9$, sendo **e)** conformação construída a apartir da estrutura

cristalográfica de referência; **f)** rotâmero gg-gg interagindo com o Ca^{2+} principal pela Man-4; **g)** rotâmero gg-gt interagindo com o Ca^{2+} principal pela Man-8; **h)** rotâmero gt-gt interagindo com o Ca^{2+} principal pela Man-4. **i)** Numeração dos resíduos de manose da glicosilação. 64

Figura 37: Métricas usadas para estimar o ΔG relativo de interação para cada rotâmero da glicosilação $GlcNAc_2Man_5$: **a)** MM/PBSA; **b)** Igualmente de Jarzynski, as barras de erro são menores do que os símbolos do gráfico; **c)** Força de ruptura, o eixo y está invertido para facilitar a comparação com os outros gráficos. O valor obtido para o caso construído com base na estrutura cristalográfica (PDB ID 1SL4) está destacado em vermelho. 65

Figura 38: Comportamento das 165 trajetórias de simulação dirigidas feitas com v = 200 Å/ns e k = 1000 pN/Å. **a)** Perfis e distribuição de trabalho; **b)** Trabalho médio e estimativa de diferença de energia livre de acordo com a igualdade de Jarzynski. 67

Figura 39: Distribuições de trabalho calculadas de simulações de pulling com v = 200 Å/ns (preto), 100 Å/ns (vermlho), 10 Å/ns (azul) e 1 Å/ns (verde). Simulações feitas com k=166 pN/Å, k=1000 pN/Å, *ensemble* NPT e *ensemble* NVT estão mostradas na mesmas cores. 68

Figura 40: Comportamento das simulações de *pulling* feitas com v = 1 Å/ns e k = 1000 pN/Å. **a)** amostragem de 11 trajetórias ; **b)** amostragem de 123 trajetórias. 69

Figura 41: PMFs descrevendo a interação entre o CRD do DC-SIGN e a glicosilação $GlcNAc_2Man_5$ calculados com os métodos Umbrela Sampling (preto), igualdade de Jarzynski com 11 trajetórias independentes com v = 1 Å/ns e k =1000 pN/Å (azul) e igualdade de Jarzynski com 123 trajetórias independentes com v = 1 Å/ns e k =1000 pN/Å (vermelho). 70

Figura 42: Definição do volume de busca dos *dockings* executados durante o ensaio de *virtual screening*. **Linha superior:** DC-SIGN; **linha inferior:** MR. **Coluna da esquerda:** visão "frontal". **Coluna da direita:** rotação de 90º em relação à coluna da esquerda. A proteína está mostrada como *cartoon* verde, Ca^{2+} como esferas azuis e a caixa que define o volume de busca está destacada com contornos pretos. 72

Figura 43: **a)** Estrutura 2D da Kanamycin. **b)** Interação entre DC-SIGN e Kanamycin. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em sticks verdes. O ligante está mostrado em *sticks* amarelos. Os Ca^{2+} estão mostrados como esferas azuis. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 73

Figura 44: **a)** Estrutura 2D do Iohexol. **b)** Interação entre DC-SIGN e Iohexol. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 74

Figura 45: **a)** Estrutura 2D da Oxytetracycline. **b)** Interação entre DC-SIGN e Oxytetracycline. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 75

Figura 46: **a)** Estrutura 2D do Paclitaxel. **b)** Interação entre DC-SIGN e Paclitaxel. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como

esferas azuis. O ligante está mostrado em *sticks* amarelos. O resíduo Phe-313 está destacado em esferas verdes transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 75

Figura 47: **a)** Estrutura 2D do Brown HT como um sal de sódio. **b)** Interação entre DC-SIGN e Brown HT. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 76

Figura 48: **a)** Estrutura 2D da Amikacin. **b)** Interação entre MR e Amikacin. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 77

Figura 49: **a)** Estrutura 2D do Iohexol. **b)** Interação entre MR e Iohexol. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 78

Figura 50: **a)** Estrutura 2D do Chloramphenicol. **b)** Interação entre MR e Chloramphenicol. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. O resíduo Tyr-729 está destacado em esferas transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 79

Figura 51: **a)** Estrutura 2D do Ceftaroline Fosamil. **b)** Interação entre MR e Ceftaroline Fosamil. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. O resíduo Leu-694 está destacado em esferas transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto. 80

Figura 52: **a)** Estrutura 2D do Metocurine. **b)** Interação entre MR e Metocurine. O MR está mostrado em *cartoon* cinza. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. Não há ligações de hidrogênio entre o receptor e o ligante. 80

LISTA DE TABELAS

Tabela 1: Simulações de dinâmica molecular referentes aos resultados mostrados discutidos neste trabalho 28

Tabela 2: Número médio de ligações de hidrogênio estabelecidas entre DC-SIGN e cada um dos resíduos de manose dos fragmentos Man_5 e Man_9 ao longo das simulações. 51

Tabela 3: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_5 e cada resíduo do DC-SIGN. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados. 52

Tabela 4: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_9 e cada resíduo do DC-SIGN. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados. 52

Tabela 5: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com o fragmento Man_5. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho. 53

Tabela 6: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com o fragmento Man_9. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho. 53

Tabela 7: Número médio de ligações de hidrogênio estabelecidas entre MR e cada um dos resíduos de manose dos fragmentos Man_5 e Man_9 ao longo das simulações. 56

Tabela 8: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_5 e cada resíduo do MR. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados. 57

Tabela 9: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_9 e cada resíduo do MR. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados. 57

Tabela 10: Ocupância das ligações de hidrogênio de cada resíduo do MR com o fragmento Man_5. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho. 58

Tabela 11: Ocupância das ligações de hidrogênio de cada resíduo do MR com o fragmento Man_9. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho. 58

Tabela 12: Parâmetros para todas as simulações dirigidas fora do equilíbrio feitas com o sistema que descreve a interação entre o CRD do DC-SIGN e a glicosilação $GlcNAc_2Man_5$ construída conforme a estrutura cristalográfica de referência. 68

Tabela 13: Energia livre de interação entre DC-SIGN e os substratos naturais Man_5 e Man_9 e ligantes selecionados a partir do *virtual screening*. Todos os termos que compõem o valor de $\Delta G_{MM/PBSA}$ estão mostrados: interações de dispersão, interações eletrostáticas, solvatação polar e solvatação apolar. As diferenças de entropia do soluto não foram calculadas nesse caso. Portanto os resultados devem ser interpretados apenas comparativamente. 82

Tabela 14: nergia livre de interação entre MR e os substratos naturais Man_5 e Man_9 e ligantes selecionados a partir do *virtual screening*. Todos os termos que compõem o valor de $\Delta G_{MM/PBSA}$ estão mostrados: interações de dispersão, interações eletrostáticas, solvatação polar e solvatação apolar. As diferenças de entropia do soluto não foram calculadas nesse caso. Portanto os resultados devem ser interpretados apenas comparativamente. 83

LISTA DE ABREVIATURAS E SIGLAS

AMBER	*Assisted Model Building with Energy Refinement*
CD	*Cluster of Differentiation*
CLEC5A	*C-type LECtin domain family 5 member A*
CRD	*Carbohydrate Recognition Domain*
CTL	*C-Type Lectin*
CTLD	*C-Type Lectin-like Domain*
DAP12	*DNAX-Activating Protein of 12 kDa*
DC-SIGN	*Dendritic Cell-Specific ICAM3-Grabbing Non-integrin*
Fuc	Fucose
Gal	Galactose
gg	*gauche-gauche*
Glc	Glucose
GlcNAc	N-Acetil Glucosamina
GROMACS	GROningen Machine for Chemical Simulations
gt	gauche-trans
HF	Hartree-Fock
HPC2N	*High Performance Computing Center North*
ICAM3	*InterCellular Adhesion Molecule 3*
KTH	*Kungliga Tekniska Högskolan*
Man	Manose
MBP-A	*Mannose Binding Protein-A*
MDL-1	*Myeloid DAP12-associated Lectin-1*
MOPAC	*Molecular Orbital PACkage*
MR	*Mannose Receptor*
PAMP	*Pathogen-Associated Molecular Pattern*
PDB	*Protein DataBank*
PDC	*Parallel Dator Centrum*
PRR	*Pattern Recognition Receptor*
RESP	*Restrained ElectroStatic Potential*
RM1	*Recife Model 1*
tg	trans-gauche
WHAM	*Weighted Histogram Analysis Method*
ZINC	*ZINC Is Not Commercial*

SUMÁRIO

Resumo..... i
Abstract..... ii
Lista de figuras..... iii
Lista de tabelas..... ix
Lista de abreviaturas e siglas..... x

1. INTRODUÇÃO..... 1
1.1. Dengue..... 1
1.2. Estrutura e função de glicosilações..... 5
1.3. Lectinas tipo-C..... 9
1.3.1. DC-SIGN..... 9
1.3.2. MR..... 12
1.3.3. CLEC5A..... 15
1.4. Modelagem molecular..... 17
1.4.1. *Docking* e *Virtual Screening*..... 17
1.4.2. Simulações de dinâmica molecular..... 18
1.4.3. MM/PBSA..... 21
1.4.4. Força de Ruptura..... 22
1.4.5. Igualdade de Jarzynski..... 23
1.4.6. *Umbrella Sampling*..... 24

2. OBJETIVOS..... 25

3. MATERIAIS E MÉTODOS..... 25

4. RESULTADOS E DISCUSSÃO..... 30
4.1. Papel do Ca^{2+} no domínio das Lectinas Tipo-C..... 31
4.2. Mecanismos de reconhecimento CTL–*high-mannose*..... 48
4.2.1. DC-SIGN..... 49
4.2.2. MR..... 55
4.3. Rotâmeros da glicosilação *high-mannose*..... 59
4.4. O comportamento da igualdade de Jarzynski..... 66
4.5. Seleção de moléculas promissoras via ensaios de *virtual screening*... 71
4.5.1. DC-SIGN..... 73
4.5.2. MR..... 76
4.6. Refinamento de poses selecionadas do *virtual screening*..... 81
4.6.1. DC-SIGN..... 81
4.6.2. MR..... 82

5. CONCLUSÕES..... 84

6. REFERÊNCIAS..... 87

APÊNDICE A..... 99
APÊNDICE B..... 100
APÊNDICE C..... 101
APÊNDICE D..... 102
APÊNDICE E..... 103
APÊNDICE F..... 104
APÊNDICE G..... 105

1. INTRODUÇÃO

O presente trabalho envolve detalhes moleculares da interação do vírus da Dengue com receptores celulares no contexto do processo de infecção viral. Assim, nesta seção introdutória são abordados os elementos essenciais da doença "Dengue" —fatores determinantes do processo de infecção, receptores celulares e ligantes envolvidos— e todo o ferramental de modelagem molecular relevante para o presente processo investigativo.

1.1. Dengue

Dengue é uma doença viral transmitida por mosquito que atualmente ameaça aproximadamente metade da população mundial. Seus vetores, mosquitos do gênero Aedes entre os quais principalmente o *Aedes aegypt*, são bem adaptados a áreas urbanas de países tropicais e subtropicais. Tal adaptação foi historicamente facilitada pela globalização e atualmente a urbanização descontrolada e o aquecimento global contribuem para a expansão do mosquito para novas áreas[1]. A dengue representa não só um problema sanitário para os países afetados, mas também impõe um fardo social e econômico: Foi estimado, por exemplo, que o custo global da Dengue em 2011 foi de 40 bilhões de dólares americanos devido a custos diretos com tratamentos e custos indiretos devido a dias de trabalho produtivo perdidos[2] No entanto, mesmo assim a Dengue atraiu pouca atenção de indústrias farmacêuticas, sendo atualmente classificada pela Organização Mundial da Saúde como uma das 17 doenças tropicais negligenciadas[3].

O ciclo de transmissão da Dengue usualmente envolve apenas humanos e mosquitos. A fêmea do vetor usa o sangue de vertebrados como fonte de nutrientes para completar o desenvolvimento de seus ovos. Ao picar um animal infectado pelo vírus da Dengue o mosquito é infectado através do seu intestino e, após pouco mais de uma semana, o vírus acumula-se nas suas glândulas salivares. Assim, futuras picadas levarão à inoculação de vírus diretamente no sangue de seu alvo e existe possibilidade de infecção, permitindo o fechamento do ciclo[4]. A grande adaptação dos vetores às áreas urbanas torna os humanos as fontes preferenciais de sangue para o *Aedes spp*. A estratégia de prevenção da Dengue adotada, baseada no controle e eliminação dos vetores, tem como objetivo a interrupção do ciclo de infecção. Apesar da ausência de evidências científicas sobre sua efetividade, falta de otimização e desconhecimento dos riscos para a população, esse procedimento continua sendo empregado[5].

A infecção pelo vírus da Dengue pode causar um amplo espectro de manifestações clínicas. Em um dos extremos desse espectro temos infecções assintomáticas. Apesar de esse

ser um cenário desejável do ponto de vista clínico, as infecções assintomáticas representam um desafio para a vigilância epidemiológica porque são invisíveis para o sistema de saúde ao mesmo tempo em que contribuem para epidemias por ainda fazerem parte do ciclo de transmissão viral[6]. Já os casos sintomáticos vão desde sintomas inespecíficos semelhantes à gripe e fortes dores nas articulações, até casos potencialmente fatais classificados pelo Ministério da Saúde conjuntamente como “Dengue grave”. Na Dengue grave pode haver extravasamento de plasma que eventualmente leva a hemorragias massivas internas ou externas e ao quadro de choque circulatório[7].

A maior parte da patogenicidade do vírus da Dengue se deve a sua capacidade de estimular exageradamente o sistema imune do hospedeiro. A superestimulação do sistema imune do hospedeiro leva ao acúmulo de complexos virus–anticorpo nas articulações e vasos, desenvolvimento de autoimunidade contra plaquetas e células endoteliais e liberação exagerada de citocinas pró inflamatórias[8]. Todos esses mecanismos contribuem para os sintomas da Dengue grave. A existência de quatro sorotipos do vírus da Dengue é um fator complicador. O hospedeiro desenvolve imunidade contra o sorotipo específico que o infectou devido a produção de anticorpos. No entanto, eles são não-neutralizantes para os demais sorotipos e, no caso de infecção posterior por um sorotipo diferente, ocorre o fenômeno de infecção aumentada por anticorpo[9]. Como consequência, a infecção e os mecanismos de patogenicidade ocorrem de forma mais rápida e intensa, razão pela qual o risco do desenvolvimento das formas graves da doença é maior caso o paciente já tenha tido Dengue anteriormente[10]. Esse fenômeno representa um desafio para o desenvolvimento de vacinas e é razão para maior atenção epidemiológica em regiões nas quais há a presença de mais de um sorotipo simultaneamente[11].

Apesar de avanços no desenvolvimento de uma vacina, ainda há barreiras para seu uso em escala populacional nos países afetados. Anos de desenvolvimento da vacina tetravalente Dengvaxia®, do laboratório *Sanofi Pasteur*, culminaram na sua aprovação para uso em alguns países, entre os quais o Brasil em 2015. No entanto, a eficácia demonstrada pela vacina, 60% em média, é muitas vezes considerada baixa para uso em escala populacional[12]. A necessidade de três doses e o custo das vacinas dificultam ainda mais sua implementação em programas nacionais de vacinação dos pontos de vista logístico e econômico[13]. Além disso, em Dezembro de 2017 o laboratório *Sanofi Pasteur* concluiu que a vacina dobra a chance de desenvolvimento de Dengue severa em pacientes que nunca tiveram Dengue antes de serem vacinados. A Organização Mundial da Saúde atualmente recomenda que a vacina seja usada apenas em pessoas que já tiveram Dengue[14].

Não há tratamentos específicos para a Dengue. Ao ser admitido no sistema de saúde, o paciente recebe medicamentos para tratar os sintomas da doença —como febres e dores— até que seu organismo consiga eliminar o vírus naturalmente, o que geralmente acontece em menos de duas semanas após a infecção. Nos casos mais graves o paciente é acompanhado diariamente e, caso necessário, é feita internação para hidratação por via intravenosa ou recebimento de transfusões de sangue[7].

Cada uma das etapas do ciclo de replicação viral pode ser alvo para fármacos. O vírus da Dengue faz parte da família *Flaviviridae* e gênero *Flavivirus*. Trata-se de um vírus que tem uma fita simples de RNA de senso positivo, (+)ssRNA, com cerca de 11 kilobases (kb). Seu ciclo de replicação inicia-se com a adesão do vírus à célula do hospedeiro através de receptores celulares[15]. A interação vírus-receptor introduz o vírus na célula através de endocitose mediada por receptor, levando o vírus para lisossomos a fim de que seja destruído (Fig. 1). No entanto, o vírus subverte a maquinaria celular. A diminuição do pH do endossomo ativa o envelope viral, causando uma alteração conformacional que o funde à membrana do endossomo e libera o RNA viral dentro do citosol. O RNA é traduzido diretamente por ribossomos em uma única poliproteína, que é processada por proteases celulares e virais (Fig. 1). O vírus é montado conforme avança do retículo endoplasmático para o Golgi e é finalmente liberado para o ambiente extracelular. As etapas de adesão, entrada, fusão, tradução, processamento, montagem e brotamento viral podem sofrer interferência de fármacos (Fig. 1). Os alvos moleculares para fármacos podem ser elementos virais, como a glicoproteína do envelope ou a protease, ou proteínas da maquinaria celular do hospedeiro, como receptores de superfície ou enzimas da via de glicosilação[16.] O grande avanço recente na obtenção de estruturas cristalográficas das proteínas estruturais –E, M e C– e não estruturais –NS1, NS3 e NS5– do vírus permitiu um maior entendimento dos mecanismos moleculares envolvidos no ciclo de replicação do vírus, bem como estimulou iniciativas de descoberta racional de fármacos[17].

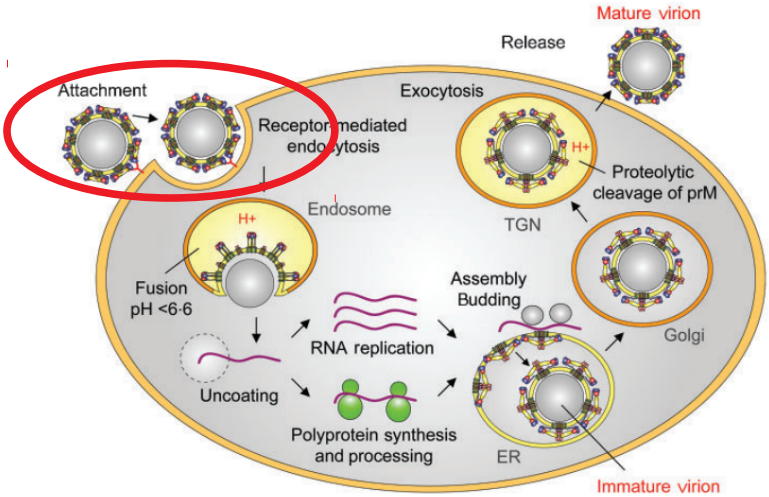

Figura 1: Representação esquemática do ciclo de replicação do vírus da Dengue. A etapa de adesão e entrada está destacada dentro da elipse vermelha. Figura adaptada da referência 15.

Estudos estruturais têm dado atenção especial às proteínas e à organização do envelope viral por estar exposto ao ambiente extracelular e às células do sistema imune. Seu elemento básico, a glicoproteína do envelope, ou proteína E, foi caracterizado por cristalografia de raios-X[18,19]. Já o método de crio eletromicroscopia, que permite a visualização em resolução quase atômica de complexos proteicos maiores do que algumas dezenas de nm (nanômetro, unidade equivalente a 10^{-9} metros), tem sido usado para investigar como essas proteínas, e outros elementos, se organizam no envelope viral[20,21]. Especificamente, o envelope é composto por 90 dímeros da proteína E em um arranjo consideravelmente flexível formando um icosaedro com aproximadamente 50nm. A flexibilidade do envelope pode ser ilustrada pelo fato de que existem diferenças estruturais à temperatura ambiente, 28ºC[22], e à temperatura interna do hospedeiro, 37ºC[23]. Além disso, a presença de anticorpos contra a glicoproteína do envelope pode alterar significativamente sua organização geral[24].

O envelope viral apresenta glicosilações que são reconhecidas por receptores celulares. Cada monômero da glicoproteína do envelope viral tem dois sítios de N-glicosilação em Asn-67 e Asn-153 (Fig. 2). Apesar de o tipo de glicosilação presente em cada um desses sítios variar com o tipo celular usado nos experimentos, há indícios de que o vírus produzido pelas células do mosquito apresentam duas glicosilações do tipo *high-mannose*. Já o vírus produzido por células humanas apresenta pelo menos uma glicosilação *complex-type* enquanto a outra varia entre *high-mannose* e *complex-type*[25-28]. Essas glicosilações são reconhecidas por receptores em células do sistema imune, como lectinas tipo C presentes em células dendríticas e macrófagos[29,30]. A interação entre o vírus e essas lectinas já foram

apontadas como alvos promissores para desenho de fármacos antivirais[31]. Portanto, no tópico seguinte serão abordados a estrutura e função de carboidratos e glicosilações.

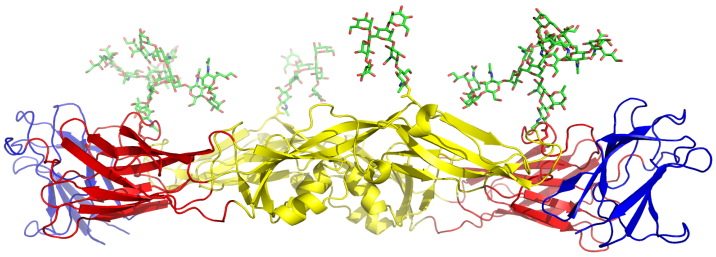

Figura 2: Dímero da glicoproteína do envelope viral. A proteína está mostrada em cartoon. Os domínios I, II e III estão mostrados em vermelho, amarelo e azul, respectivamente. As glicosilações *high-mannose* estão mostradas como *sticks* verdes.

1.2. Estrutura e função de glicosilações

Quimicamente, carboidratos são poli-álcoois com uma função cetona ou aldeído. Dependendo do tamanho da cadeia, a carbonila do grupo aldeído ou cetona pode sofrer reação intramolecular com uma das hidroxilas, formando um hemicetal ou hemiacetal heterocíclico. As formas fechadas –isto é, cíclicas– mais comuns são as que tem seis ou cinco átomos como parte do anel e são chamadas de piranoses e furanoses, respectivamente. O carbono da função hemicetal ou hemiacetal é chamado anomérico porque as duas possíveis orientações da sua hidroxila na forma fechada criam dois isômeros chamados de anômeros. Quando o oxigênio da hidroxila do carbono anomérico está mais distante do oxigênio endocíclico o anômero é chamado de α e, de forma inversa, quando esse oxigênio está mais próximo do oxigênio endocíclico o anômero é chamado de β. Anômeros α e β e as formas piranose e furanose se interconvertem em solução porque todas coexistem em equilíbrio tendo a forma aberta como intermediário comum[32].

Enzimas polimerizam carboidratos de forma análoga a amino ácidos. Através de um processo enzimático, o grupo hemiacetal ou hemicetal de um monossacarídeo pode reagir com uma das hidroxilas de outro, formando uma ligação glicosídica que resulta em um dissacarídeo e uma molécula de água. Uma vez que a reação pode acontecer em mais de uma hidroxila de um dos substratos, é necessário especificar qual delas está envolvida na ligação glicosídica. Por convenção, os carbonos dos monossacarídeos são numerados a partir da extremidade da molécula que tem o carbono anomérico e as hidroxilas recebem os mesmos números do carbono ao qual estão ligadas. Por exemplo, a nomenclatura do dissacarídeo lactose é βGal(1,4)Glc: um monossacarídeo de Galactose (Gal) ligado através da forma β de seu hemiacetal, que está no carbono 1, à hidroxila 4 de um monossacarídeo de Glucose (Glc). Ao contrário dos demais biopolímeros encontrados na natureza—DNA, RNA e proteínas—o

processo de polimerização de carboidratos pode gerar moléculas ramificadas, ou seja, um mesmo resíduo de carboidrato pode estar envolvido em mais de duas ligações glicosídicas[32].

A ligação do tipo α1,6 é diferente das demais devido a presença de uma ligação rotacionável extra. O ângulo diedral que descreve a rotação ao dessa ligação, ω, é definido como O6-C6-C5-O5. As três conformações metaestáveis do ângulo ω são chamadas de *gauche-gauche* (gg), *gauche-trans* (gt) e *trans-gauche* (tg) e correspondem ao valor do ângulo ω de -60º (ou 300º), 60º e 180º, respectivamente. Os nomes dos rotâmeros de referem à orientação (*trans* ou *gauche*) em relação aos ângulos diedros ω e O6-C6-C5-C4, respectivamente (Fig. 3) [33,34].

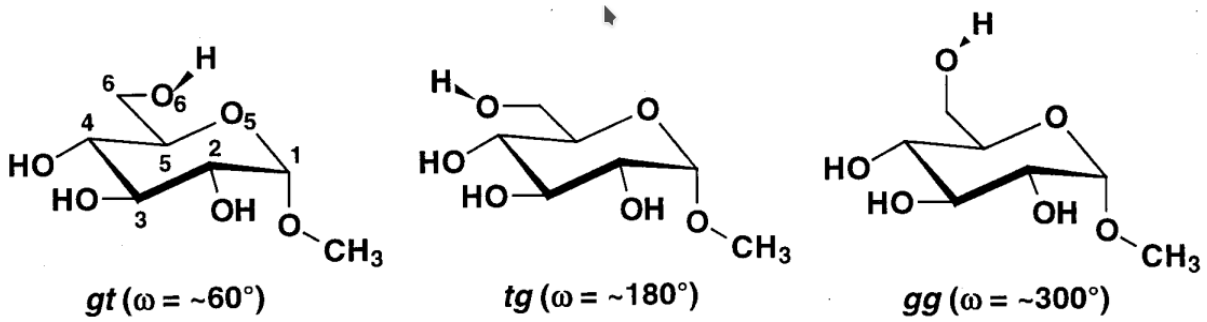

Figura 3: Rotâmeros da ligação do tipo α1,6. Figura extraída da referência 33.

Mecânica molecular não é suficiente para explicar a preferência conformacional de algumas ligações glicosídicas. Além das considerações convencionais feitas no contexto da mecânica molecular, um efeito estereoeletrônico é relevante para o entendimento da estabilidade relativa dos rotâmeros gg, gt e tg: Um dos pares de elétrons solitários (do inglês "*lone pair*") do oxigênio endocíclico, que corresponde a um orbital não-ligante chamado n, pode interagir com o orbital anti-ligante não ocupado da ligação covalente entre C1 e o oxigênio exocíclico que participa da ligação glicosídica, chamado σ*. Os rotâmeros *gauche* teriam menor estabilidade do que o *trans* de acordo com considerações estéricas simples. No entanto, a orientação dos orbitais n e σ* permite que eles formem um orbital molecular estendido que aumenta sua estabilidade. Devido a esse fenômeno os rotâmeros gg e gt tem preferência sobre o rotâmero tg[35-38].

Carboidratos podem ser covalentemente ligados a proteínas. Além da reação de polimerização entre dois carboidratos, eles também podem ser covalentemente ligados a proteínas através de dois tipos de processo enzimático: N-glicosilação ou O-glicosilação. Na O-glicosilação ocorre polimerização de monossacarídeos com o átomo de oxigênio de cadeias laterais de Serina, Threonina, Tirosina, Hidroxilisina e Hidroxiprolina. A O-glicosilação pode, a partir dai, crescer conforme novos monômeros são adicionados a esse monossacarídeo inicial devido a ação de enzimas. Já no processo de N-glicosilação ocorre polimerização entre

um oligossacarídeo previamente construído e o nitrogênio de cadeias laterais de Asparagina ou Treonina[32]. O processo de glicosilação efetivamente aumenta a diversidade do proteoma e exerce funções diversas como contatos célula–célula, contatos célula–matriz extracelular e pode auxiliar o processo de enovelamento de proteínas[39].

O processo de N-glicosilação ocorre no lúmen do retículo endoplasmático (Fig. 4). Inicialmente um carboidrato precursor é adicionado à proteína e processado em uma glicosilação que tem nove resíduos de Manose (Man) e dois de N-AcetilGlucosamina (GlcNAc), $GlcNAc_2Man_9$. Conforme a proteína glicosilada avança do retículo plasmático para a rede cis-Golgi ocorre o processo de *trimming*, no qual enzimas presentes no retículo endoplasmático podem remover resíduos de Manose que fazem ligação do tipo α1-2 no oligossacarídeo inicial. Uma vez que o processamento de glicosilações é dinâmico e não está sujeito a controles rígidos na célula, todas as moléculas de $GlcNAc_2Man_9$ a $GlcNAc_2Man_5$ podem existir simultaneamente e são coletivamente chamadas de glicosilação do tipo *high-mannose*[32]. Do ponto de vista estrutural, os casos extremos —$GlcNAc_2Man_5$ e $GlcNAc_2Man_9$—são inequivocamente definidos porque nenhuma ou todas as Manoses que fazem ligações α1-2 foram removidas enquanto os casos intermediários —$GlcNAc_2Man_8$ a $GlcNAc_2Man_6$—são heterogêneos porque há várias formas possíveis de remover resíduos de Manose da glicosilação inicial[32]. Adicionalmente, do ponto de vista conformacional, já que todas as glicosilações *high-mannose* tem duas ligações do tipo α1-6, há quatro possíveis combinações de rotâmeros mais estáveis: gg-gg, gg-gt, gt-gg e gt-gt[40,41].

Glicosilações *high-mannose* podem ser processadas conforme a proteína avança para a rede trans-Golgi (Fig. 4). Já na parte medial do Golgi pode ocorrer adição de moléculas de Glucose à glicosilação *high-mannose*, dando origem ao tipo de glicosilação conhecido como híbrida. Além disso, o *trimming* das Manoses pode continuar e outros açúcares podem ser ligados à Glucose, criando as glicosilações *complex-type*. Trata-se de um grupo grande e diverso de glicosilações que exerce várias funções, geralmente como sinalização celular através de contatos célula–célula e célula–matriz, estabilização de proteínas e controle fino da atividade enzimática. O aparato enzimático necessário para processar as glicosilações *high-mannose* completamente até *complex-type* está ausente ou é incompleto em patógenos como vírus, bactérias, fungos e protozoários[32]. Por isso, enquanto células humanas têm majoritariamente glicosilações *complex-type* em sua superfície, glicosilações *high-mannose* são reconhecidas como um Padrão Molecular Associado a Patógeno (PAMP, do inglês "*Pathogen-Associated Molecular Pattern*") por células do sistema imune. O reconhecimento desses PAMPs por Receptores Reconhecedores de Padrões (PRRs, do inglês "*Pattern Recognition Receptors*") ativa imediatamente as vias bioquímicas da imunidade inata,

causando inflamação e início da resposta imune adaptativa para geração de anticorpos contra o patógeno específico[42,43].

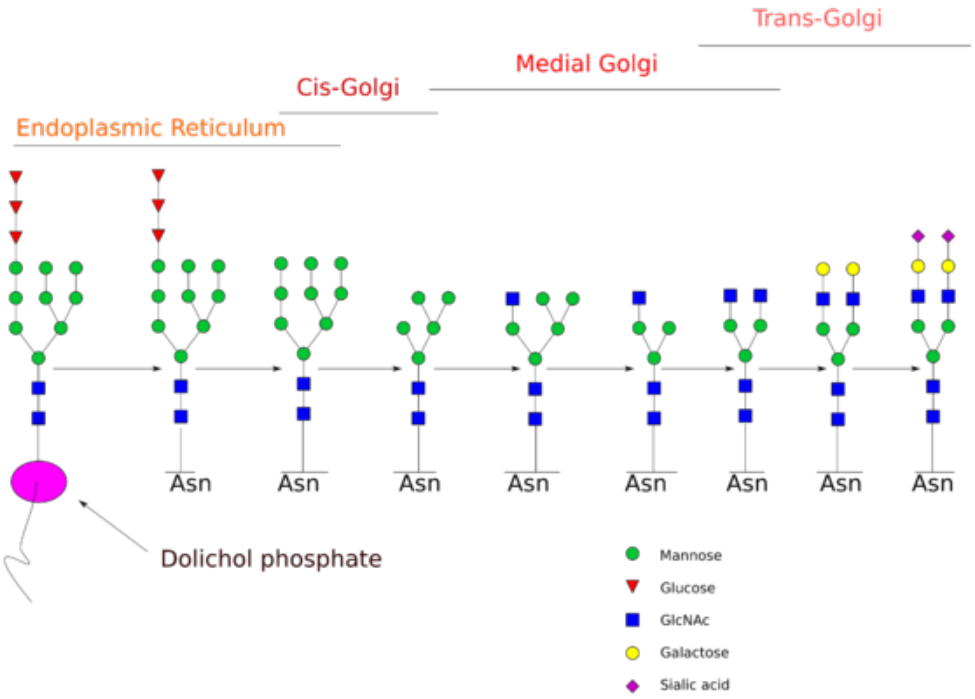

Figura 4: Representação esquemática do processo de *trimming* e processamento da glicosilação *high-mannose* inicial, no retículo endoplasmático, até a *complex-type* no trans-Golgi.

Carboidratos são as macromoléculas biológicas com maior potencial de armazenamento de informação na natureza. Eles superam o potencial de armazenamento de informação dos ácidos nucleicos e amino ácidos devido à grande variabilidade de formas através das quais dois monômeros podem se ligar para formar ligações glicosídicas, incluindo a possibilidade de ramificações[32].

A variedade de monômeros disponíveis é muito maior nos carboidratos. Assim, monossacarídeos formam glicanas da mesma forma que letras formam palavras numa língua que é chamada de "glico código"[44]. Proteínas especializadas em reconhecer carboidratos, chamadas genericamente de "lectinas", fazem a leitura do glico código e transduzem a informação química presente nas glicosilações em função biológica através de uma série de mecanismos específicos de cada grupo de lectinas[45]. Em particular, Lectinas Tipo-C (CTL, do inglês "*C-Type Lectins*") reconhecem as glicosilações presentes no envelope do vírus da Dengue e contribuem para a patogenicidade da doença[46]. Detalhes sobre são discutidos a seguir.

1.3. Lectinas tipo-C

A família das Lectinas Tipo-C (CTL, do inglês "*C-Type Lectins*") é a maior e mais diversa família de lectinas presente em animais. Seu nome, "Tipo-C", vem de sua característica principal de ter um domínio de reconhecimento de carboidratos (CRD, do inglês "*Carbohydrate Recognition Domain*") que complexa até quatro Ca^{2+} e que tem atividade lectínica Ca^{2+}-dependente. Atualmente considera-se que a família CTL é parte da superfamília de proteínas que tem Domínios de Lectina Tipo-C (CTLD, do inglês "*C-Type Lectin-like Domain*"). Os membros da superfamília CTLD estão subdivididos em 17 grupos e não necessariamente complexam Ca^{2+} ou carboidratos apesar de terem o domínio CTL. Em células da linhagem mielóide, como neutrófilos, macrófagos e células dendríticas, são encontrados apenas os grupos II, V e VI[47-49].

Neste trabalho foi considerado um representante de cada um desses grupos que tem relevância na infecção pelo vírus da Dengue, a saber: DC-SIGN (do inglês, "*Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin*"), CLEC5A (do inglês "*C-type LECtin domain family 5 member A*") e MR (do inglês "*Mannose Receptor*"), respectivamente.

1.3.1. DC-SIGN

O receptor DC-SIGN (do inglês, "*Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin*"), também chamado de CD209 (do inglês, "*Cluster of Differentiation 209*"), é expresso em células dendríticas e está envolvido em processos de adesão celular, regulação da resposta imune e reconhecimento de moléculas endógenas e exógenas[50]. Ele se organiza em *nanoclusters* na superfície celular e, após ser estimulado pela complexação de seus ligantes, sofre internalização rápida por mecanismo dependente de clatrina. O complexo DC-SIGN–ligante é levado para compartimentos proteolíticos onde o ligante, possivelmente um patógeno como o vírus da dengue, é processado para ser apresentado a linfócitos T no processo chamado de "apresentação de antígenos" que é crucial para a continuidade da resposta imune. É interessante notar que esse receptor passa por reciclagem; após acidificação da vesícula para onde o complexo DC-SIGN–ligante vai, o receptor perde afinidade pelos seus ligantes, que são liberados, enquanto o receptor volta íntegro para a parte externa da membrana celular[51].

Além de sua atividade endocítica, DC-SIGN tem a capacidade de induzir diferentes padrões de resposta imune de acordo com os carboidratos presentes nas estruturas que reconhece[52]. Em particular, ao reconhecer glicosilações do tipo *high-mannose*, sua porção intracelular desencadeia cascatas de sinalização celular que resultam na produção de citocinas

pró-inflamatórias através da ativação Raf-1 e NF-κB[53]. Por outro lado, o reconhecimento de antígenos que apresentam glicosilação *complex-type* com Fucose terminal suprime a resposta inflamatória através do desmonte do complexo de sinalização intracelular na cauda citoplasmática do DC-SIGN[54]. Esse receptor também está envolvido no desenvolvimento de processos alérgicos[55].

Foi observado que o DC-SIGN é o principal receptor usado pelo vírus da Dengue para infectar células dendríticas, que são seu alvo primário. Células THP-1, que não expressam DC-SIGN e não são infectadas pelo vírus da Dengue, foram usadas para validar o papel específico do receptor na infecção. O gene do DC-SIGN foi transfectado para essas células, que passaram a expressá-lo e a serem infectadas. A quantidade de células infectadas é proporcional à quantidade de DC-SIGN presente na superfície das células THP-1 transfectadas. Anticorpos anti-DC-SIGN foram capazes de diminuir a infecção das células THP-1 transfectadas e de células dendríticas imaturas em até 90%[29]. Apesar de ainda haver dúvidas a respeito do mecanismo exato através do qual DC-SIGN leva à internalização do vírus[56,57], o papel essencial e específico das glicosilações do envelope viral no processo de entrada do vírus em células que expressam DC-SIGN foi demonstrado pelo fato de que vírus mutantes que não apresentam glicosilações não conseguem infectar tais células[58] e pela capacidade de agentes complexadores de carboidratos inibirem a infecção de células que expressam DC-SIGN[59]. Além de seu envolvimento na Dengue, o receptor DC-SIGN está envolvido nas infecções por outros patógenos. Por isso, inibidores do DC-SIGN tem sido procurados na literatura[60] usando abordagens como glicodendrímeros[61,62], glicomiméticos[63,64] e pequenas moléculas sintéticas[65,66].

Há bastante informação estrutural sobre o receptor DC-SIGN devido ao grande interesse terapêutico sobre ele. O receptor completo tem 404 resíduos de amino ácido, sendo 37 da cauda citoplasmática, 21 em uma hélice transmembrana, 198 divididos em oito domínios repetidos chamados de “pescoço” e, finalmente, o domínio de reconhecimento de carboidratos (CRD, do inglês “*Carbohydrate Recognition Domain*”) com 124 resíduos e uma porção C-terminal desorganizada com 24 resíduos (Fig. 5a). Na superfície celular o receptor é encontrado como tetrâmeros estabilizados pelas repetições das hélices do pescoço[67,68], que formam uma estrutura do tipo haste rígida ligada de forma flexível aos CRDs[69]. Uma consequência de tal flexibilidade entre os CRDs e o pescoço é que, conforme revelado por uma estrutura obtida por crio-microscopia eletrônica, apenas dois CRDs do receptor tetramérico interagem simultaneamente com as glicosilações do envelope viral[70]. A estrutura a baixa resolução de toda a porção extracelular do receptor foi modelada através do uso combinado de espalhamento de raios-x a baixo ângulo e da avaliação do seu comportamento hidrodinâmico[71,72] (Fig. 5b). Adicionalmente, foi demonstrado que a tetramerização das repetições do domínio do pescoço é dependente de pH, o que pode ser o mecanismo

responsável pela perda de afinidade entre o receptor e seus ligantes ao ocorrer acidificação das vesículas após endocitose[72].

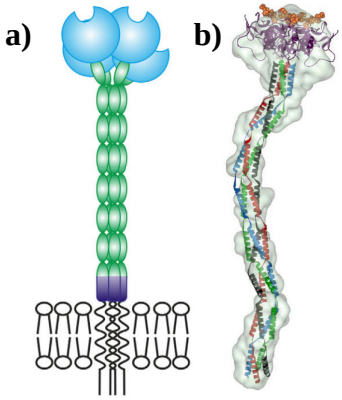

Figura 5: a) Representação esquemática da estrutura completa do DC-SIGN. Os CRDs estão representados em azul e os domínios do pescoço em verde. Figura adaptada da referência 68. **b)** Estrutura 3D à baixa resolução da porção extracelular do DC-SIGN. Figura adaptada da referência 72.

A estrutura do seu CRD é composta pelo *fold* conservado das CTL complexando o total de três Ca^{2+}, sendo um deles no sítio principal, que está envolvido na complexação de açúcares, e dois deles em um mesmo sítio, chamado de auxiliar[73] (Fig. 6a). Os resíduos diretamente envolvidos na complexação do Ca^{2+} são Glu-347, Asn-349, Glu-354, Asn-365 e Asp-366. Os mesmos resíduos formam uma rede de ligações de hidrogênio com um dos resíduos de manose do carboidrato que está complexado ao receptor, que é chamada de “manose principal”. O DC-SIGN tem um sítio de reconhecimento de carboidratos bastante extenso e que pode reconhecer vários padrões de carboidratos[74].

Experimentos de g*lycan array*, nos quais a afinidade do receptor por centenas de padrões de carboidrato diferentes é avaliada simultaneamente, elucidaram as propriedades de reconhecimento molecular do DC-SIGN. Além da afinidade relativa por diferentes glicosilações do tipo *high-mannose*, foi detectada ligação específica a glicosilações *complex-type* com fucose terminal como os antígenos sanguíneos A e B e Lewis a, b, x e y[75]. Um par de estruturas cristalográficas do DC-SIGN complexando dois tetrassacarídeos –um fragmento da glicosilação *high-mannose* (PDB ID 1SL4) (Fig. 6b) e outro que contém o padrão *complex-type* Lewix x (PDB 1SL5) (Fig. 6c)– revelam os detalhes dessas interações[75]. Estruturas cristalográficas de complexos entre DC-SIGN e inibidores glicomiméticos mostram que os mecanismos de interação receptor-carboidrato podem ser explorados por inibidores[76,77].

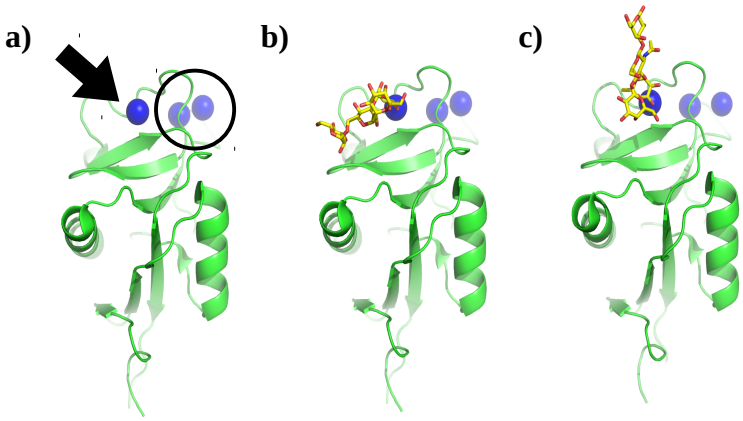

Figura 6: Estruturas cristalográficas do DC-SIGN. A proteína está mostrada em cartoons verdes, os Ca^{2+} como esferas azuis e os carboidratos como *sticks* amarelos. **a)** O sítio de complexação principal de Ca^{2+} está destacado pela seta preta. O sítio auxiliar, que complexa dois Ca^{2+}, está destacado dentro do círculo preto; **b)** Estrutura do complexo com o fragmento Man_4 (PDB ID 1SL4) **c)** Estrutura do complexo com o fragmento Le^x (PDB ID 1SL5).

É interessante notar que as glicosilações que o DC-SIGN reconhece mantêm-se consideravelmente flexíveis mesmo após a formação do complexo. Em uma estrutura cristalográfica obtida usando um fragmento grande da glicosilação *high-mannose*, contendo seis resíduos de manose, foi possível resolver a densidade eletrônica de apenas três deles. Isso é compatível com valores altos para o fator de temperatura nos resíduos de carboidrato que não estão diretamente ligados ao Ca^{2+} principal em outras estruturas, indicando movimento ou heterogeneidade estrutural nesses pontos da estrutura dos açúcares complexados. A existência de heterogeneidade estrutural é reforçada pela observação, por cristalografia de raios-x, de que é possível haver mais de um modo de interação dos carboidratos no sítio de complexação do DC-SIGN[78].

1.3.2. MR

O Receptor de Manose (MR, do inglês "*Mannose Receptor*"), também conhecido como CD206, é membro do grupo VI da superfamília CTLD, caracterizado por ter vários CRDs com atividade lectínica dependente de Ca^{2+}. Esse receptor é expresso em macrófagos, células dendríticas imaturas e alguns tipos de células endoteliais[79] e está envolvido na remoção de glicoproteínas endógenas potencialmente danosas da circulação e no reconhecimento, fagocitose e processamento de ligantes destinados ao processo de apresentação de antígenos[80]. Assim como o DC-SIGN, esse receptor endocítico também passa por reciclagem, isto é, ao se envolver no processo de endocitose ele perde afinidade por seu ligante em pH ácido e retorna para a membrana celular, onde volta a reconhecer seus ligantes[81].

Foi demonstrado que o MR é essencial para a infecção produtiva de macrófagos pelo vírus da Dengue e que a interação entre o receptor e o vírus da Dengue ocorre através das glicosilações do envelope viral. Células 3T3, que não expressam MR e não são infectadas pelo vírus da dengue, tornam-se suscetíveis à infecção após serem transfectadas com o gene do MR humano. Além disso, anticorpos anti-MR diminuem a produção de vírus e o número de macrófagos infectados em aproximadamente 80%[30]. Ligantes com alta afinidade especificamente pelo MR tem sido buscados para fazer a entrega de moléculas diretamente para células apresentadoras de antígeno a fim de desenvolver vacinas mais eficientes e seguras[82], entrega de fármacos a células específicas[83] e realização de exames de imagem[84].

O receptor completo é formado por um domínio rico em cisteínas, um domínio de fibronectina tipo II, oito CTLDs em sequência, uma hélice transmembrana e uma região intracelular para sinalização (Fig. 7a). Através de ensaios de ligação a açúcares usando formas truncadas do receptor foi determinado que os CTLDs 1–3 tem afinidade baixa por carboidratos e que os CTLDs 4–8 são necessários e suficientes para a atividade lectínica. Dentro desse fragmento, CTLD 4 é o único domínio que tem afinidade relevante por monossacarídeos de Manose, Fucose, N-Acetil Glucosamina e Glucose quando isolado dos demais. Embora o CTLD 4 isolado tenha qualitativamente a mesma afinidade por carboidratos do que o receptor inteiro, a presença dos CTLDs 5 e 7 são necessárias para atingir a alta afinidade do receptor completo, provavelmente porque estabelecem interações adicionais com os ligantes multivalentes e porque orientam os domínios da forma correta[85].

De forma geral o mecanismo de complexação de carboidratos no CTLD 4 do MR é similar ao observado para o DC-SIGN. A atividade lectínica é dependente de Ca^{2+} e do pH em ambos porque resíduos ao redor do Ca^{2+} principal participam da interação com os açúcares, incluindo coordenação direta do Ca^{2+} pelas hidroxilas do carboidrato[86]. Em pH ácido, como das vesículas para onde o receptor vai junto com seus ligantes, há perda de afinidade pelo Ca^{2+} devido a protonação de resíduos ácidos que o complexam. Isso causa perda da afinidade pelo ligante e uma mudança conformacional na qual o CTLD 4, inicialmente ativo e resistente à ação de proteases na presença de Ca^{2+}, torna-se inativo e é lisado por proteases em pH ácido. Apesar de evidências indiretas de que o CTLD 4 do MR complexe dois Ca^{2+}, alinhamentos de sequência indicam que os sítios de complexação de Ca^{2+} auxiliares presentes em outros membros da família estão ausentes no MR[87]. A localização de um possível sítio auxiliar de Ca^{2+} no MR foi avaliada através de mutagênese e de um modelo estrutural construído com base numa proteína com 28% de identidade, a MBP-A (do inglês "*Mannose Binding Protein-A*") de rato. No entanto, foi concluído que, caso o MR realmente ligue dois Ca^{2+}, o sítio através do qual no cátion adicional é complexo é único dentro da família[88].

A estrutura 3D do CTLD 4, que corresponde aos resíduos 626–768 da proteína completa, foi determinada por cristalografia de raios-X (PDB IDs 1EGG e 1EGI). Apesar de estas estruturas cristalográficas terem dado uma importante contribuição para o entendimento conformacional das CTLs, há um inconveniente importante: Ocorreu troca de domínios (do inglês, "*domain swap*") entre membros de duas unidades assimétricas diferentes no cristal. Especificamente, os resíduos 708–728 da proteína em uma unidade assimétrica se projetam para longe do centro da estrutura e participam da composição do domínio de seu vizinho de simetria (Fig. 7b). Embora esse fenômeno seja biologicamente relevante para o mecanismo de ação de algumas proteínas[89], trata-se de um artefato induzido pelas condições de cristalização uma vez que o CTLD 4 do MR é monomérico em solução. Sua estrutura geral, incluindo o sítio de complexação do Ca^{2+} principal, é muito similar a outras CTL como DC-SIGN. Os resíduos que complexam o Ca^{2+} principal são Glu-725, Asn-727, Asn-728, Asn-747 e Asp-748. O possível sítio de complexação do Ca^{2+} auxiliar, predito com base em evidências indiretas, não foi observado. No entanto, sua ausência nas estruturas cristalográficas não exclui totalmente a possibilidade de sua existência uma vez que a troca de domínios pode ter afetado a organização da região onde ele estaria[90]. Recentemente estruturas cristalográficas dos quatro domínios N-terminais do MR, a saber o domínio rico em cisteínas, fibranectina tipo II e CTLDs 1–2, mostrou que o CTLD-1 não complexa Ca^{2+} e que o CTLD-2 complexa apenas o Ca^{2+} principal[91].

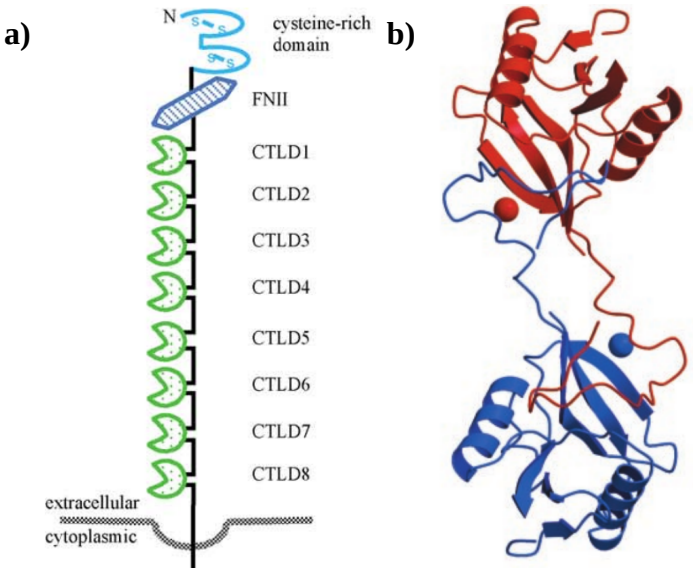

Figura 7: a) Representação esquemática da estrutura completa do MR. Adaptado da referência 79. **b)** Estrutura cristalográfica do CTLD-4 do MR no qual aconteceu troca de domínios entre unidades assimétricas vizinhas no cristal, que estão representados em cores diferentes. Figura adaptada da referência 90.

1.3.3. CLEC5A

O receptor CLEC5A (do inglês "*C-type LECtin domain family 5 member A*"), também conhecido como MDL-1 (do inglês, "*Myeloid DAP12-associated Lectin-1*"), é membro do grupo V da superfamília CTLD, caracterizado por ter um domínio do tipo CTLD que não tem atividade lectínica e não liga Ca^{2+}. Esse receptor é expresso em várias populações de células do sistema imune, como células dendríticas imaturas, macrófagos e neutrófilos. Ao interagir com seus ligantes, como o vírus da Dengue, o receptor CLEC5A inicia cascatas de sinalização intracelular através de sua interação com a proteína adaptadora DAP12 (do inglês, "*DNAX-Activating Protein of 12 kDa*") que resultam na produção de citocinas que amplificam a resposta imune[92].

Foi demonstrado que CLEC5A interage com o vírus da Dengue, aumentando a adesão de vírus a células 293T transfectadas com o gene do receptor. Manana, um polímero grande e heterogêneo composto por Manose, e o monossacarídeo Fucose diminuem a adesão do vírus a essas células em até 50%. Ao contrário do DC-SIGN e MR, a interação CLEC5A–vírus não induz endocitose e infecção, mas sim a ativação de cascatas de sinalização intracelular, iniciadas pela fosforilação de DAP12, que resultam na liberação de citocinas pró-inflamatórias. Em macrófagos, tanto o silenciamento do gene que codifica o receptor CLEC5A quanto o uso de anticorpos anti-CLEC5A diminuem a fosforilação de DAP12 e liberação de citocinas pró-inflamatórias sem causar impacto na replicação viral ou quantidade de células infectadas. Uma característica interessante da inibição do CLEC5A é que secreção de Interferon-α, uma citocina que estimula a resposta imune contra vírus, não é afetada. *In vivo*, esses anticorpos diminuem o extravasamento de líquido e hemorragia causados pela infecção do vírus da Dengue em camundongos imunocomprometidos. Enquanto 100% desses camundongos imunocomprometidos morrem 14 dias após a infecção se não forem tratados, o uso de anti-CLEC5A leva a uma taxa de sobrevivência de 50% e permite a eliminação total do vírus da circulação em 23 dias[93]. Apesar do reconhecimento do potencial terapêutico da inibição do CLEC5A para alívio dos sintomas de Dengue severa[94,95] e de complicações de outras infecções como pelos vírus Influenza[96] e Vírus da Encefalite Japonesa[97], ainda não tentativas publicadas de encontrar pequenas moléculas que possam ser usadas como inibidores.

A despeito das importantes consequências da interação entre CLEC5A e o vírus da Dengue, a afinidade entre eles é baixa[98]. Recentemente foi proposto que o CLEC5A interage com vírus previamente aderidos à célula através de receptores de alta afinidade, como MR e DC-SIGN. Nesse contexto, foi demonstrado que o bloqueio da interação entre o vírus e MR ou DC-SIGN é suficiente para inibir a amplificação da resposta imune iniciada pelo CLEC5A[99]. Além disso, não se conhece o mecanismo da interação entre CLEC5A e o

vírus da Dengue; apesar de evidências indicando que o reconhecimento se daria através das glicosilações do envelope, não foi detectada afinidade da CLEC5A por nenhum carboidrato de dois painéis contendo representantes de glicosilações comuns na natureza, incluindo as presentes no vírus da Dengue[100].

O receptor completo tem uma sequência curta de resíduos, uma hélice transmembrana, um domínio *linker* e o domínio CTLD (Fig. 8a), cuja estrutura 3D foi resolvida por cristalografia de raios-X (PDB ID 2YHF). Na superfície celular o CLEC5A forma homodímeros independentemente da presença de DAP12 e há evidências de que a dimerização ocorra pelo domínio *linker* possivelmente através de formação de pontes dissulfeto entre duas cadeias. A estrutura do seu CTLD segue o *fold* conservado da superfamília (Fig. 8b), sendo que as maiores diferenças entre CLEC5A e os receptores DC-SIGN e MR estão, conforme esperado, concentradas na região que está envolvida com complexação de Ca^{2+} e carboidratos nos outros receptores. Tal região, chamada de "*loop* maior" na família CTLD, é mais curta no CLEC5A: contém 12 resíduos, entre 149 e 160, em comparação com 20 no DC-SIGN, entre 336 e 355. Dois resíduos ácidos do sítio de complexação de Ca^{2+} conservado na família são hidrofóbicos no CLEC5A, explicando a sua falta de afinidade por esse cátion. Além disso, enquanto não há interações via cadeia principal entre o *loop* maior e o restante da estrutura no DC-SIGN, no CLEC5A dois resíduos estão envolvidos na formação de folha β com o núcleo da estrutura[100].

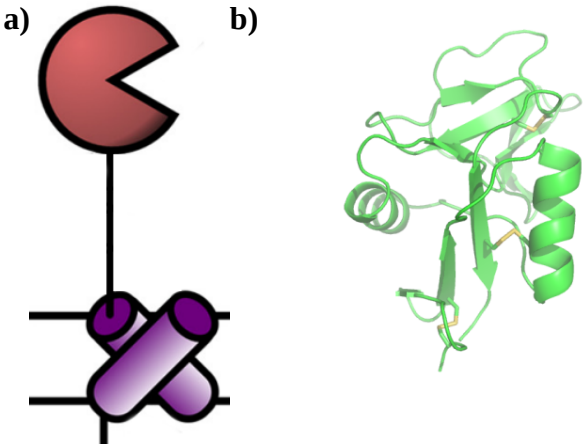

Figura 8: a) Representação esquemática da estrutura completa do CLEC5A. Seu CLTD está mostrado em vermelho, o domínio linker como uma linha e sua hélice transmembrana está em roxo. A interação com DAP12 ocorre dentro da membrana. Figura adaptada da referência 100. **b)** Estrutura cristalográfica do seu CTLD (PDB ID 2YHF).

1.4. Modelagem molecular

Uma grande variedade de métodos computacionais pode ser classificado como modelagem molecular. A seguir descrevemos aqueles que foram aplicados ao longo desta tese.

1.4.1. *Docking* e *Virtual Screening*

Docking, ou docagem molecular, é um método que tem como objetivo determinar a melhor forma através da qual um complexo receptor–ligante pode se formar a partir das estruturas do receptor e do ligante isolados. Apesar da grande variedade de algoritmos e programas disponíveis, a base do método é a geração de complexos candidatos que obedecem restrições estéricas. Isso é feito através de algoritmos de busca que variam a conformação do ligante e a atribuição de um valor de *score* a cada uma dessas "poses" — nome dado às estruturas geradas via *docking*. Ao fim da busca uma seleção de poses representativas com bons valores de *score* e modos de interação diversos são reportados através de análise estatística[101].

Idealmente, os valores de *score* teriam alta correlação com a afinidade observada experimentalmente e conseguem diferenciar claramente ligantes que tem alta afinidade pelo receptor daqueles que tem baixa afinidade. No entanto, enquanto os algoritmos usados atualmente conseguem gerar poses que reproduzem bem resultados experimentais, as funções de *score* ainda são um fator limitante, sendo comum a ausência de correlação entre *score* e afinidade experimental ou a presença de resultados falsos-positivos, ou seja, bons valores de *score* serem atribuídos a poses diferentes do que seria experimentalmente observado ou poses de ligantes que não tem afinidade pelo receptor serem reportadas como tendo bons valores de *score*[102].

Quando o *docking* é usado de forma massiva através de bibliotecas que contém milhares ou até milhões de estruturas de moléculas candidatas a ligantes, o método passa a ser chamado de *virtual screening* ou triagem virtual. Embora sofra das mesmas limitações do *docking* em relação à função de *score* usada, atualmente a triagem virtual é ferramenta corriqueira na academia e na indústria nas etapas iniciais de iniciativas de descoberta ou desenho de fármacos[103]. É considerado uma forma rápida e barata de reduzir o número de moléculas a serem testadas experimentalmente de milhões para centenas ou até dezenas[104].

1.4.2. Simulações de dinâmica molecular

Um dos métodos mais versáteis dentro da modelagem molecular é a simulação de dinâmica molecular, que essencialmente requer três ingredientes para obter seu resultado final: *i*) estrutura 3D da macromolécula de interesse, geralmente obtida por cristalografia de raios-X ou ressonância magnética nuclear; *ii*) campo de forças, constituído por um conjunto de equações e parâmetros previamente calibrados para reproduzir alguma referência como dados experimentais ou cálculos quânticos; *iii*) instruções específicas de simulação e análise, que detalham quais procedimentos o programa de simulação deve executar, quais condições físico-químicas considerar e qual tipo de informação extrair dos resultados brutos, chamados de trajetória[105].

Dessa forma, torna-se possível usar a estrutura inicial para resolver as equações do movimento de Newton iterativamente segundo as energias calculadas pelo campo de forças para descrever a evolução conformacional da macromolécula a partir de seus movimentos em escala atômica frente a diferentes condições ou estímulos. Simulações de dinâmica molecular foram usadas para, por exemplo, identificar mecanismos alostéricos em proteínas[106], verificar alterações conformacionais induzidas por ligantes[107], avaliar efeitos do *crowding* macromolecular[108], investigar propriedades mecânicas e dinâmicas de vírus inteiros[109], caracterizar mecanismos de interação receptor–ligante[110] e descrever o funcionamento de canais transmembrana na presença da bicamada lipídica[111]. Simulações de dinâmica molecular também são usadas em conjunto com outras técnicas de modelagem para, por exemplo, refinamento de modelos 3D obtidos por algoritmos de predição de estrutura[112] e refinamento de poses *docking*[113].

O cerne do método de simulação de dinâmica molecular, que é exposto a seguir, tem-se mostrado robusto apesar de constantes inovações nos modelos e algoritmos disponíveis[114,115]: Um volume de simulação é definido ao redor da estrutura 3D inicial do soluto, ao qual são adicionadas moléculas de água para considerar os efeitos do solvente e íons em quantidade suficiente para neutralizar as cargas do soluto e atingir a força iônica desejada. A partir daí é possível alimentar as equações do campo de forças escolhido com as coordenadas iniciais de cada átomo que compõe o sistema para calcular a energia potencial de cada um deles. De forma geral, a energia potencial total do sistema, U, é a soma das contribuições covalentes e não-covalentes (Eq. 1). Tradicionalmente, enquanto a contribuição covalente é constituída pela energia devido ao estiramento de ligações covalentes, dobramento de ângulos planares e torção de ângulos diedros (Eq. 2), a contribuição não-covalente é a soma das interações eletrostática e de dispersão entre os átomos (Eq. 3).

$$U = U_{covalente} + U_{não-covalente} \tag{1}$$

$$U_{covalente} = U_{ligações} + U_{ângulos} + U_{diedros} \tag{2}$$

$$U_{não-covalente} = U_{eletrostática} + U_{dispersão} \tag{3}$$

Mais especificamente, a energia potencial proveniente do estiramento de ligações é modelada por um potencial harmônico que descreve como o comprimento de cada ligação, b, está deformado em relação ao seu comprimento de repouso, b_0. A constante de mola k_b fornece a escala energética para tal estiramento (Eq. 4).

$$U_{ligações} = \sum_{ligações} \frac{1}{2} k_b (b - b_0)^2 \tag{4}$$

De forma análoga, o dobramento de ângulos planares é modelado por um potencial harmônico que descreve como cada ângulo planar, θ, está distorcido em relação a seu valor de repouso, θ_0. A escala energética é dada pela constante de mola k_θ (Eq. 5).

$$U_{ângulos} = \sum_{ângulos} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{5}$$

Já a torção dos ângulos diedros ϕ é modelada por funções periódicas com diferentes multiplicidades n e ângulo de deslocamento de fase ϕ_s. A escala energética é dada pelas constantes k_ϕ (Eq. 6).

$$U_{diedros} = \sum_{diedros} \frac{1}{2} k_\phi \left[1 + \cos(n\phi - \phi_s) \right] \tag{6}$$

No caso das contribuições não-covalentes, as interações eletrostáticas entre cada par de átomos i e j é modelada pelo potencial de Coulomb de acordo com a permissividade elétrica do vácuo, ϵ_0, a carga parcial de cada átomo, q_i e q_j, e a distância entre eles, r_{ij} (Eq. 7). A contribuição das interações de dispersão, presentes mesmo na ausência completa de carga, são modeladas pelo potencial de Lennard-Jones, que tem como parâmetros a energia de interação ϵ entre o par de átomos e a distância na qual a energia de interação σ entre eles é zero. A energia varia em função da distância r_{ij} entre os átomos (Eq. 8).

$$U_{eletrostática} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{7}$$

$$U_{dispersão} = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{8}$$

A partir das equações previamente descritas, que modelam a energia potencial do sistema em função das coordenadas atômicas, a força F_i que atua sobre cada átomo i pode ser calculada como o negativo do gradiente espacial da energia potencial (Eq. 9) e, a

partir da força, a aceleração a_i que cada átomo experimenta naquele instante do tempo é calculada (Eq. 10). O último passo do algoritmo básico da dinâmica molecular é atualizar a posição de cada átomo segundo um pequeno intervalo de tempo Δt pré-determinado, que é da escala de femtossegundos, isso é, 10^{-15} segundos. Assim, cada átomo tem sua posição e velocidade atualizadas através das equações do movimento de Newton de acordo com a aceleração que está experimentando no instante atual da simulação. Por fim, após a posição e velocidade de cada átomo ter sido atualizada, o tempo atual da simulação é incrementado Δt (Eq. 11) e o ciclo se fecha; a energia potencial é recalculada com as novas coordenadas e velocidades para dar prosseguimento à construção da trajetória. A simulação termina após um número pré-determinado de passos, usualmente na escala de milhões, e dezenas de simulações podem ser necessárias para que se consiga um único resultado pontual.

$$\vec{F}_i = -\nabla U \tag{9}$$

$$\vec{a}_i = \frac{d^2 \vec{r}_i}{d t^2} = \frac{\vec{F}_i}{m_i} \tag{10}$$

$$t_{atual} = t_{anterior} + \Delta t \tag{11}$$

Apesar do funcionamento básico do algoritmo ser essencialmente o descrito acima, há muitos outros aspectos na implementação dos programas de dinâmica molecular que contribuem para a robustez do método, como controle de temperatura e pressão, otimizações nos cálculos para diminuir o tempo de execução, paralelização para permitir uso simultâneo de grande quantidades de núcleos de processamento e desenvolvimento de modelos mais complexos para melhorar a qualidade dos resultados[114,115].

Uma quantidade central para o entendimento do comportamento de uma ampla variedade de sistemas é a diferença de energia livre, ΔG, entre dois estados porque revela a probabilidade de transição entre eles. Ao definirmos o estado final como "complexado" — aquele no qual receptor e ligante interagem formando um complexo— e o estado inicial como "dissociado" —no qual receptor e ligante estão separados e não há interação significativa entre eles— o valor de $\Delta G_{\text{interação}}$ reflete a afinidade entre receptor e ligante. Há vários métodos para fazer esse cálculo, que frequentemente é um dos computacionalmente mais custosos entre as aplicações de simulações de dinâmica molecular[116]. A seguir discutimos alguns desses métodos.

1.4.3. MM/PBSA

O método MM/PBSA (do inglês, "*Molecular Mechanics / Poisson-Boltzmann Surface Area*") usa informações apenas dos estados complexado e dissociado para calcular a diferença de energia livre entre eles. Isso é feito através da soma de duas contribuições energéticas (Eq.12): *i*) interações moleculares entre receptor e ligante e *ii*) diferenças de solvatação entre receptor e ligante complexados e livres (Eq. 13). As contribuições da interação direta entre receptor e ligante são calculadas no vácuo como a diferença de entalpia e de entropia vibracional entre os estados complexado e dissociado de acordo com os modelos de mecânica molecular convencionais (Eq. 14)[117].

Por outro lado, as diferenças de energia livre de solvatação para uma dada molécula ou complexo são calculadas como a soma das contribuições da polar, que corresponde às interações eletrostáticas soluto–solvente, e apolar, correspondendo à energia necessária para abrir uma cavidade no solvente e às interações de dispersão soluto–solvente (Eq. 15). Enquanto o termo de solvatação polar é calculado resolvendo-se numericamente e equação de Poisson-Boltzmann para modelar a transição da molécula ou complexo do vácuo, modelado como um meio contínuo cuja constante dielétrica ε é 1, para a água, modelado como um solvente contínuo com ε=80 (Eq.16). O termo apolar é calculado como uma relação linear empírica com a área de superfície acessível ao solvente, *A*, através do termo de tensão superficial γ e um termo de correção *c* (Eq. 17)[118]. A relação entre todos os termos e $\Delta G_{interação}$ pode ser visualizada através de um ciclo termodinâmico (Fig. 9).

$$\Delta G_{interação} = \Delta G^{vácuo}_{interação} + \Delta G_{solvatação} \tag{12}$$

$$\Delta G_{solvatação} = \Delta G^{complexo}_{solvatação} - \left(\Delta G^{receptor}_{solvatação} + \Delta G^{ligante}_{solvatação}\right) \tag{13}$$

$$\Delta G^{vácuo}_{interação} = \Delta H^{vácuo} - T\,\Delta S^{vácuo}_{vibracional} \tag{14}$$

$$\Delta G_{solvatação} = \Delta G^{polar}_{solvatação} + \Delta G^{apolar}_{solvatação} \tag{15}$$

$$\Delta G^{polar}_{solvatação} = \left(G_{\epsilon=80} - G_{\epsilon=1}\right) \tag{16}$$

$$\Delta G^{apolar}_{solvatação} = \gamma\, A_{exposta} + c \tag{17}$$

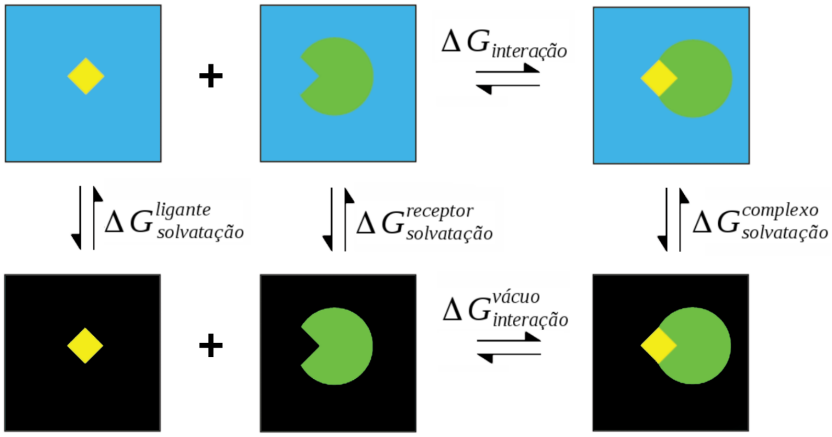

Figura 9: Ciclo termodinâmico esquemático para o método MM/PBSA. O solvente está representado pelo fundo azul e o vácuo pelo fundo preto. Figura adaptada de "*AMBER Advanced Tutorials – Tutorial 3*".

Trajetórias obtidas por simulações de dinâmica molecular podem ser pós-processadas usando o método MM/PBSA. Isso permite o cálculo de diferenças de energia livre de interação a partir de uma amostragem conformacional detalhada, melhorando a convergência e qualidade dos resultados. Devido a limitações do método MM/PBSA —como não considerar o efeito da participação de moléculas de água como pontes para interações receptor–ligante e complicações na escolha ou interpretação de parâmetros devido ao uso misto de modelagem molecular e eletrostática contínua— os valores de energia livre calculados geralmente não são absolutos, isto é, não podem ser comparados diretamente com as afinidades receptor–ligante observadas experimentalmente[119]. Mesmo assim, a combinação de simulações de dinâmica molecular e MM/PBSA tem sido aplicada com sucesso a casos de desenho e descoberta de fármacos devido a sua versatilidade, precisão e custo computacional relativamente baixo para calcular valores de $\Delta G_{interação}$ relativos, isso é, que indicam corretamente a ordem de afinidade de uma série de ligantes por um receptor[120] e que descrevem a contribuição de cada resíduo do receptor para a afinidade por seus ligantes[121].

1.4.4. Força de Ruptura

Simulações de dinâmica molecular dirigidas (do inglês, "*Steered Molecular Dynamics*") fora do equilíbrio podem ser usadas para definir uma mola virtual, isso é, um potencial harmônico, que conecta o receptor e o ligante e cujo comprimento de repouso aumenta durante a simulação a uma taxa constante. Conforme o comprimento de equilíbrio da mola aumenta ele torna-se maior do que a distância de equilíbrio entre receptor e ligante. Isso leva ao acúmulo de força na mola até o ponto em que as interações moleculares são rompidas. A força de ruptura, definida como a força máxima que a mola impõe sobre o complexo durante a simulação dirigida (Fig. 10), tem sido usada recentemente como um estimador de afinidade relativa altamente correlacionado com $\Delta G_{interação}$ experimentalmente observado[122-124].

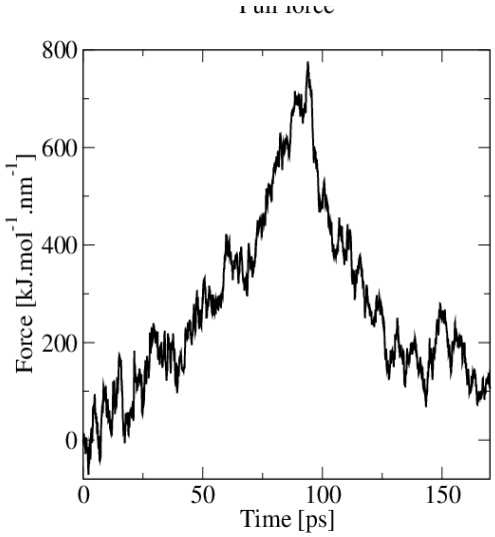

Figura 10: Força que a mola virtual impõe sobre o sistema receptor–ligante para separá-los durante uma trajetória representativa de simulação de dinâmica molecular dirigida. A força aumenta ao longo to tempo e alcança um máximo, que é a força de ruptura, antes de romper as interações receptor–ligante.

1.4.5. Igualdade de Jarzynski

Além da força de ruptura, simulações dirigidas permitem que calculemos o trabalho feito pela mola sobre o sistema. No limite de baixa velocidade de *pulling*, ou seja, alongamento da mola lento, tem-se um processo no equilíbrio e o trabalho feito pela mola sobre o sistema é igual a $-\Delta G_{interação}$, onde o sinal negativo está presente porque durante a simulação dirigida estamos indo do estado final, complexado, para o inicial, dissociado. No entanto, tal processo lento demandaria muito tempo de simulação. Em aplicações reais a mola alonga-se rápido demais e o processo ocorre fora do equilíbrio. Como consequência, os valores de trabalho calculados a partir dessas trajetórias podem ser mais negativos do que $-\Delta G_{interação}$ porque há dissipação de energia durante a simulação[125,126].

Ao executarmos várias realizações independentes da simulação dirigida que separa receptor e ligante podemos calcular o trabalho feito pela mola em cada uma delas e, assim, construir uma distribuição de trabalho fora do equilíbrio. De acordo com a Igualdade de Jarzynski é possível usar a distribuição de trabalho fora do equilíbrio para obter $\Delta G_{interação}$ de equilíbrio desde que um número suficientemente grande de realizações independentes seja feita [127] (Eq.7).

$$e^{-\beta \Delta G} = \langle e^{-\beta W} \rangle \tag{7}$$

Apesar de a validade da Igualdade de Jarzynski já ter sido verificada experimentalmente[128], ainda há poucas aplicações computacionais para cálculo de $\Delta G_{interação}$ em sistemas de interesse biológico[129]. Isso possivelmente ocorre porque o número de simulações dirigidas independentes que são necessárias para calcular $\Delta G_{interação}$ com acurácia é dependente da velocidade de *pulling*. Dessa forma, o custo computacional para aplicá-la pode

ser proibitivamente alto; foi reportado, por exemplo, que o uso da Igualdade de Jarzynski para calcular o perfil de energia livre de ligantes atravessando uma proteína de canal implica custo computacional 10 vezes maior do que *Umbrella Sampling*, um método já bem estabelecido[130].

1.4.6. *Umbrella Sampling*

Outro método que utiliza simulações dirigidas para calcular $\Delta G_{interação}$ é conhecido como *Umbrella Sampling*. Inicialmente são feitas simulações dirigidas para separar o complexo receptor–ligante a fim de construir o "caminho de reação" ao longo de uma "coordenada de reação" espacial, que geralmente é a distância entre receptor e ligante definida de forma específica para cada sistema (Fig. 11, parte superior). A partir daí, pontos ao longo do caminho de reação, que representam o sistema receptor–ligante com diferentes distâncias de separação, são selecionados e amostrados em simulações independentes nas quais potenciais harmônicos são usados para manter a coordenada de reação espacial dentro de uma série de janelas discretas (Fig. 11, parte central), isso é, simulações dirigidas nas quais o comprimento de repouso da mola é constante para cada janela[131]. A diferença de energia livre é calculada através do método de análise de histogramas ponderados (WHAM, do inglês "*Weighted Histogram Analysis Method*")[132], que calcula a distribuição de probabilidade ao longo da coordenada de reação espacial considerando o efeito do potencial harmônio, chamado de potencial influenciador (do inglês, "*biasing potential*"), aplicado para restringir a amostragem às janelas discretas[133] (Fig. 11, parte inferior).

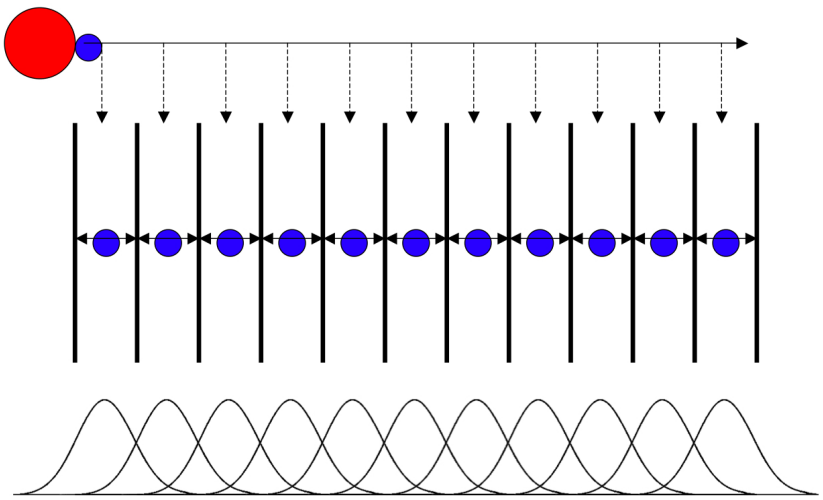

Figura 11: Representação esquemática do método *Umbrella Sampling*. **Linha superior:** Amostragem inicial da coordenada de reação através de simulações dirigidas. **Linha central:** Simulações de amostragem individuais nas quais são aplicados potenciais harmônicos para restringir a amostragem do caminho de reação dentro de cada janela. **Linha inferior:** Histogramas de amostragem com boa sobreposição. Fonte: "*GROMACS Tutorial – Umbrella Sampling*".

2. OBJETIVOS

O objetivo deste trabalho é a aplicação, desenvolvimento e domínio de métodos e ferramentas computacionais para:

- *i*) descrever os mecanismos moleculares do reconhecimento da glicosilação *high-mannose*, presente no envelope do vírus da Dengue, pelas lectinas DC-SIGN e MR;

- *ii*) encontrar moléculas que podem contribuir para a descoberta de *leads* contra a infecção do vírus da Dengue.

3. MATERIAIS E MÉTODOS

Estruturas iniciais

O CRD do DC-SIGN foi simulado de acordo com suas estruturas cristalográficas em complexo com os fragmentos Man_4 (PDB ID 1SL4)[73] e Le^x (PDB ID 1SL5)[73] . A estrutura sem ligante foi construída removendo-se o carboidrato da estrutura PDB ID 1SL4. Da mesma forma, diferentes estados de ligação a Ca^{2+} foram obtidos removendo-se o Ca^{2+} principal, os auxiliares ou todos os Ca^{2+} da estrutura PDB ID 1SL4.

As estruturas cristalográficas disponíveis para o CTLD-4 do MR (PDB IDs 1EGG, 1EGI)[90] apresentaram troca de domínio devido às condições de cristalização. Por isso, a estrutura de referência para as simulações foi construída de acordo com o sugerido na literatura usando a MBP-A como um molde[88,90]. Primeiro, a cadeia A da estrutura PDB ID 1EGG foi duplicada, criando as cadeias 1EGG:A1 e 1EGG:A2. A cadeia 1EGG:A1 foi sobreposta ao CRD da MBP-A (PDB ID 1MSB)[134] e o *loop* da 1EGG:A2 que sofreu troca de domínios (resíduos 707-731) foi sobreposto ao *loop* correspondente na MBP-A. A estrutura final foi montada partir dos resíduos 628-702 da 1EGG:A1, 707-727 da 1EGG:A2 e 735-763 da 1EGG:A1. Os resíduos para os quais não havia estrutura, 703-706, foram adicionados diretamente com o *software* PyMOL e os resíduos 728-734 (Asn-Tyr-Gln-Asn-Val-Glu-Tyr) foram construídos a partir do segmento Asp188-Asp194 da MBP-A, que foi mutado para coincidir com a sequência do MR. O estado livre de Ca^{2+} foi construído removendo-se o Ca^{2+} principal. As estruturas do MR em complexo com ligantes foram obtidas através de sobreposição com a estrutura do DC-SIGN complexando Man_4 (PDB ID 1SL4).

No caso do CLTD da CLEC5A, a estrutura cristalográfica PDB ID 2YHF[100] foi usada.

Os complexos do DC-SIGN e MR com os açúcares Man_5, Man_9, $GlcNAc_2Man_5$ e $GlcNAc_2Man_9$ foram construídos através da sobreposição da estrutura de cada açúcar ao fragmento Man_4 presente na estrutura cristalográfica PDB ID 1SL4. No caso dos complexos

entre as lectinas e os possíveis rotâmeros das glicosilações, cada manose de cada rotâmero estável foi sobreposta ao à manose que interage diretamente com o Ca^{2+} principal do fragmento presente na estrutura cristalográfica PDB ID 1SL4, conforme descrito na seção de resultados.

As estruturas 3D otimizadas de todos os açúcares foram obtidas pelo servidor GLYCAM-Web e a dos ligantes simulados para refinamento dos ensaios de *virtual screening* foram obtidas diretamente da biblioteca de moléculas usada no *docking*.

Parametrização dos ligantes

A parametrização dos ligantes para as simulações de dinâmica molecular foi feita com programas do pacote AmberTools[135]. As cargas parciais foram determinadas com o método RESP (do inglês, *"Restrained ElectroStatic Potential"*) a partir da densidade eletrônica calculada em nível de teoria HF/6-31G* com o *software* Gaussian. A fim de acelerar a convergência dos cálculos de otimização feitos no Gaussian executamos uma pré-otimização usando a Hamiltoniana semi-empírica RM1 (do inglês, *"Recife Model 1"*)[136] com o programa MOPAC[137].

Simulações de dinâmica molecular

As simulações de dinâmica molecular foram feitas com o pacote GROMACS[138,139], utilizando o campo de forças AMBER99SB-ILDN[140] para modelar a proteína, íons e água, GLYCAM06[141] para carboidratos e GAFF[142] para ligantes. O modelo de água usado foi o TIP3P[143]. Foram usadas condições periódicas de contorno em uma caixa dodecaédrica dentro da qual a força iônica foi mantida em 150 mM pela adição de NaCl ou $CaCl_2$ (Tabela 1). Além disso, foram adicionados contra-íons para neutralizar as cargas do sistema. O tamanho do passo foi 2 fs com as ligações covalentes da água e das demais moléculas tratadas pelos algoritmos LINCS[144] e SETTLE[145], respectivamente. Todas as interações intermoleculares foram tratadas explicitamente dentro de um raio de corte de 1 nm e as interações eletrostáticas de longo alcance foram tratadas com PME[146].

O primeiro passo dado em todas as simulações realizadas foi a minimização de energia usando o algoritmo *steepest descent*. Durante a fase de equilibração, a temperatura foi mantida em 309 K pelo termostato *v-rescale*[147], e a pressão foi mantida em 1 bar pelo barostato de Berendsen[148]. Durante a fase de produção a temperatura e pressão foram controladas pelo termostato de Nosé-Hoover[149,150] e barostato de Parrinello-Rahman[151]. Réplicas de simulações partiram da mesma estrutura inicial com diferentes velocidades iniciais atribuídas aos átomos do sistema. Nos casos em que várias simulações dirigidas foram

executadas, também usamos diferentes conformações iniciais extraídas de uma simulação de dinâmica molecular prévia.

A quantidade de réplicas e tempo de simulação variou de acordo com a aplicação específica ao longo deste trabalho (Tabela 1). De forma geral as análises foram feitas desprezando-se metade de cada trajetória para garantir equilibração do sistema e o erro estatístico foi calculado como desvio padrão da média considerando as réplicas.

MM/PBSA

No caso das estimativas de afinidade dos rotâmeros da glicosilação pelo DC-SIGN foi usado o pacote AmberTools[135]. Topologias de acordo com o formato do AMBER foram geradas com o programa tLEaP e os cálculos foram feitos com o *script* MMPBSA.py[152]. A constante dielétrica interna usada foi 4. Para os cálculos, foi extraído um total de 50 *frames* de trajetórias de simulações dos complexos lectina–açúcar. Erros foram calculados através de 200 reamostragens pelo método *bootstrap*. As diferenças de entropia do soluto foram calculadas com os mesmos 50 *frames*. Inicialmente *frame* foi submetido à otimização em precisão dupla no vácuo sem *constraints*. Então, a matriz Hessiana foi calculada e diagonalizada, gerando os autovalores e autofrequências. A entropia vibracional do complexo, do receptor e do ligante foi calculada a partir das autofrequências de todos os seus modos normais de vibração. Não houve frequências negativas e os primeiros seis autovalores não foram numericamente relevantes[153,154].

Nos casos do cálculo das contribuições do cátion e dos resíduos dos receptores para a energia livre de interação com os açúcares, foi usado o *software* g_mmpbsa[155]. Um total de 400 *frames* foram extraídos de trajetórias de dinâmica molecular e as diferenças de entropia foram ignoradas uma vez só havia interesse nas contribuições relativas.

O mesmo *software*, g_mmpbsa[155], foi usado no refinamento do *virtual screening*. Foram usados 50 *frames* extraídos de simulações de dinâmica molecular. As diferenças de entropia não foram calculadas uma vez que o objetivo era apenas obter um *ranking* das moléculas.

Os arquivos mdp, que contém todos os parametros usados para a execução dos passos de dinâmica molecular com o software GROMACS estão no fim deste documento como apêndices.

Tabela 1: Simulações de dinâmica molecular referentes aos resultados mostrados discutidos neste trabalho

Receptor	Ligante	Condição	Réplicas
DC-SIGN	-	3 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
DC-SIGN	-	2 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
DC-SIGN	-	1 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
DC-SIGN	-	0 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
DC-SIGN	-	3 Ca^{2+} complexados, $CaCl_2$ na solução	3 x 200 ns
DC-SIGN	-	0 Ca^{2+} complexados, $CaCl_2$ na solução	3 x 200 ns
MR	-	1 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
MR	-	0 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
MR	-	1 Ca^{2+} complexados, $CaCl_2$ na solução	3 x 200 ns
MR	-	0 Ca^{2+} complexados, $CaCl_2$ na solução	3 x 200 ns
CLEC5A	-	0 Ca^{2+} complexados, NaCl na solução	3 x 200 ns
CLEC5A	-	0 Ca^{2+} complexados, $CaCl_2$ na solução	3 x 200 ns
DC-SIGN	-	Força de ruptura dos Ca^{2+} complexados	378 x 0.1 ns
DC-SIGN	Man_4	Amostragem conformacional	5 x 100 ns
DC-SIGN	Le^x	Amostragem conformacional	5 x 100 ns
MR	Man_4	Amostragem conformacional	5 x 100 ns
DC-SIGN	Man_5	Amostragem conformacional	5 x 50 ns
DC-SIGN	Man_9	Amostragem conformacional	5 x 50 ns
MR	Man_5	Amostragem conformacional	5 x 50 ns
MR	Man_9	Amostragem conformacional	5 x 50 ns
DC-SIGN	$GlcNAc_2Man_5$	*Umbrella Sampling*	40 x 50 ns
DC-SIGN	$GlcNAc_2Man_9$	*Umbrella Sampling*	40 x 50 ns
DC-SIGN	Le^x	*Umbrella Sampling*	30 x 50 ns
DC-SIGN	$GlcNAc_2Man_5$	Amostragem de 11 conformações	11 x 5 x 50 ns
DC-SIGN	$GlcNAc_2Man_9$	Amostragem de 14 conformações	14 x 5 x 50 ns
DC-SIGN	$GlcNAc_2Man_5$	Igualdade de Jarzynski, v=200 Å/ns	660 x 0.15 ns
DC-SIGN	$GlcNAc_2Man_5$	Igualdade de Jarzynski, v=100 Å/ns	660 x 0.30 ns
DC-SIGN	$GlcNAc_2Man_5$	Igualdade de Jarzynski, v=10 Å/ns	220 x 3 ns
DC-SIGN	$GlcNAc_2Man_5$	Igualdade de Jarzynski, v=1 Å/ns	156 x 30 ns
DC-SIGN	4 ligantes	Refinamento do *Virtual Screening*	4 x 5 x 20 ns
MR	4 ligantes	Refinamento do *Virtual Screening*	4 x 5 x 20 ns

Docking e virtual screening

Os *dockings* individuais foram executados com o programa AutoDock Vina[156] com valor do parâmetro de exaustividade igual a 10 e volume de busca envolvendo o sítio de complexação de carboidratos. A biblioteca de ligantes foi obtida no banco de dados ZINC (“*ZINC Is Not Commercial*”)[157] a partir de catálogos de moléculas já aprovadas como drogas pelo FDA. O *virtual screening* foi automatizado, analisado e filtrado com um conjunto de *scripts* e programas de implementação própria, conforme descrito na seção de resultados e discussão.

Umbrella Sampling

A coordenada de reação espacial foi inicialmente amostrada por cinco réplicas de simulações dirigidas com velocidade de *pulling* de 10 Å/ns e constante de mola de 500 pN/Å. A constante de mola usada para as simulações de amostragem foi 166 pN/Å. A fim de evitar a interação do ligante com sítios irrelevantes para o fenômeno de interesse foi necessário aplicar restrições de posição cilíndricas tanto ao ligante quanto ao receptor. O movimento dos seus átomos foi restrito ao interior de cilindros de raio 0.5nm e 1.0nm, respectivamente. Nos cálculos de energia livre foi usado o WHAM (do inglês, “*Weighted Histogram Analysis Method*”)[158] considerando o tempo de autocorrelação integrado para eliminar dados correlacionados. O caminho de reação foi dividido em 80 janelas na análise e foram feitas 200 re-amostragens por *bootstrap* para calcular o erro estatístico.

Hardware

Ao longo do desenvolvimento deste trabalho utilizamos o seguinte *hardware*:

– Supercomputador Beskow, composto por 1676 nós com 16 cores @ 2.3 GHz cada. É administrado pelo PDC (do sueco, “*Parallel Dator Centrum*”) do KTH (do sueco, “*Kungliga Tekniska Högskolan*”), em Estocolmo, Suécia.

– Supercomputador Kebnekaise, composto por 544 nós com 28 cores @ 2.60 GHz cada e alguns com 2 ou 4 GPUs nVIDIA K80. É administrado pelo HPC2N (do inglês, “*High Performance Computing Center North*”) em Umeå, Suécia.

– Supercomputador DAVinCI, composto por 12 nós com 12 cores @ 2.80 GHz e 2 GPUs nVIDIA Tesla M2050 cada. É administrado pela Rice University, em Houston, Estados Unidos.

– Supercomputador BlueGene/Q, composto por 32 nós com 512 núcleos PowerPC @ 1.6 GHz cada. Faz parte do Convênio de Cooperação Científica entre a Rice University (Texas-EUA) e a USP.

– Infraestrutura de computação em nuvem da USP. Nosso uso máximo foi de 20 máquinas virtuais com 8 núcleos @ 2.40 GHz cada.

– *Desktops* locais com 6 núcleos @ 3.00 GHz cada para execução de passos preparatórios e análises.

4. RESULTADOS E DISCUSSÃO

O sítio de complexação principal de Ca^{2+} do DC-SIGN e do MR são estruturalmente similares e estão envolvidos no reconhecimento de açúcares. Para maior facilidade de leitura, daqui em diante iremos chamá-los de "sítio principal". De forma análoga, o sítio de reconhecimento de Ca^{2+} auxiliar do DC-SIGN, que complexa dois Ca^{2+}, será chamado simplesmente de "sítio auxiliar". Finalmente, o Ca^{2+} complexado ao sítio principal será chamado de "Ca^{2+} principal" e aqueles complexados ao sítio auxiliar de "Ca^{2+} auxiliares" (Fig. 12).

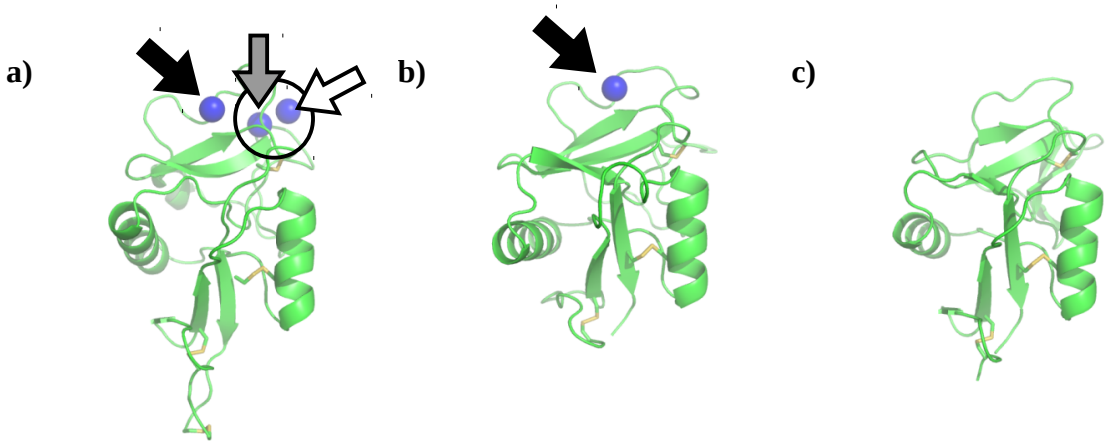

Figura 12: Estruturas de referência dos três receptores considerados: **a)** DC-SIGN; **b)** MR; **c)** CLEC5A. A proteína está mostrada em cartoon verde e os Ca2+ como esferas azuis. O Ca2+ principal está destacado por uma seta preta. O sítio de complexação auxiliar de Ca2+ do DC-SIGN está destacado dentro de um círculo preto. Os dois Ca2+ presentes no sítio auxiliar do DC-SIGN podem ainda ser individualmente identificados como "auxiliar interno" e "auxiliar externo", destacados por uma seta cinza e branca, respectivamente.

4.1. Papel do Ca^{2+} no domínio das Lectinas Tipo-C

A interação entre os três receptores considerados neste estudo e Ca^{2+}, assim como a eventual presença de sítios de complexação de Ca^{2+} não evidentes em suas estruturas cristalográficas, foram inicialmente investigados através da remoção de todos os Ca^{2+} da estrutura inicial do DC-SIGN e MR e pelo uso de $CaCl_2$ para neutralizar cargas e atingir a força iônica desejada para a caixa de simulação. A partir dai foram feitas simulações de dinâmica molecular longas para amostrar a distribuição dos cátions ao redor dos receptores. CLEC5A também foi simulado nas mesmas condições indicadas.

Observamos que durante as simulações os resíduos que compõem os sítios principal e auxiliar do DC-SIGN e MR interagem fortemente com alguns Ca^{2+} inicialmente livres na solução. Os demais resíduos desses receptores interagem com Ca^{2+} da solução apenas de forma transiente ou com resíduos ácidos isolados que não compõem um sítio de complexação (Fig. 13). O mesmo resultado foi obtido através de simulações nas quais os Ca^{2+} cristalográficos foram mantidos durante a simulação na presença de $CaCl_2$ na solução.

Esses resultados indicam que essa abordagem é eficiente para detecção de sítios de complexação de Ca^{2+} em proteínas e que não há outros sítios além dos já descritos no DC-SIGN e MR. Isso é particularmente relevante para o MR: apesar de sua estrutura cristalográfica mostrar apenas um Ca^{2+} complexado ao sítio principal de seu CTLD-4, há evidências indiretas de que poderia haver um sítio auxiliar que seria diferente de todos os outros membros da família. Não foi possível encontrar tal sítio para MR apesar de nossas tentativas de compô-lo ativamente a partir dos dados disponíveis da literatura através de uma combinação de simulações de dinâmica molecular convencionais, dirigidas e com restrições tanto no vácuo quanto em solvente explícito. Portanto, propomos que MR se liga a apenas um Ca^{2+} no seu sítio principal e que não há outros sítios de complexação a Ca^{2+} no seu domínio CTLD-4.

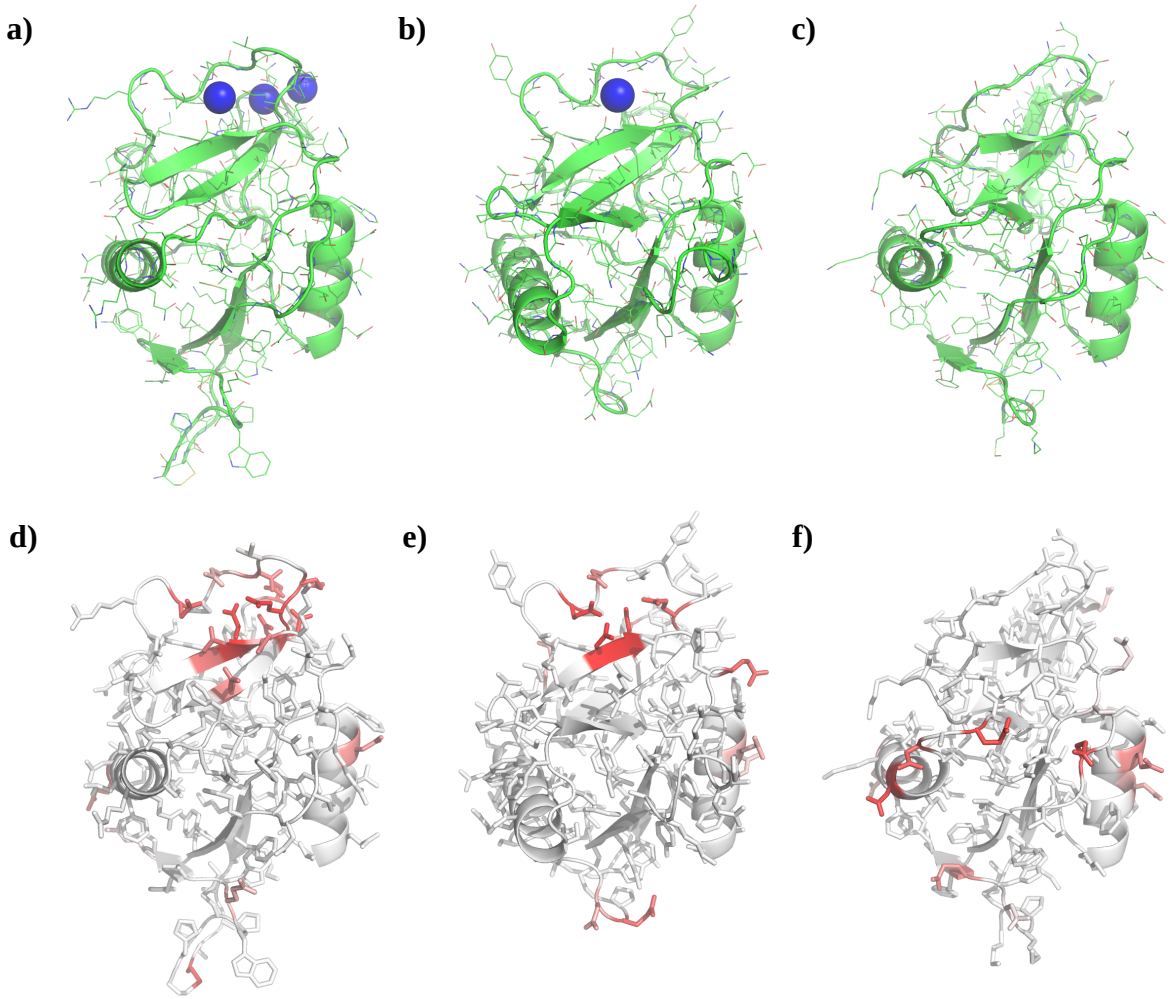

Figura 13: Interação de Ca^{2+} da solução durante simulações nas quais os Ca^{2+} cristalográficos foram removidos e $CaCl_2$ foi usado para neutralizar as cargas e ajustar a força iônica em 150 mM. **(a,d)** DC-SIGN; **(b,e)** MR; e **(c,f)** CLEC5A. **Linha superior:** estruturas cristalográficas, incluindo os Ca^{2+}, representados por esferas azuis. **Linha inferior:** ocupância da interação de cada resíduo com Ca^{2+} da solução projetada sobre a estrutura cristalográfica. A escala de cores vai de branco representando 0% de ocupância —ausência total de interação entre o resíduo e Ca^{2+}— até vermelho representando 100% de ocupância —interação do resíduo com Ca^{2+} durante todo o tempo de simulação analisado.

Interações dos receptores com Ca^{2+} e Na^{+} contribuem significativamente para a estabilidade entálpica do complexo

A fim de investigar os mecanismos moleculares da interação entre Ca^{2+} e os CTLDs de cada receptor —DC-SIGN, MR e CLEC5A— simulações de dinâmica molecular exaustivas foram feitas considerando seus estados de ligação a Ca^{2+} mais relevantes. NaCl foi usado para neutralizar as cargas do sistema e atingir a força iônica fisiológica. Três contribuições distintas para a energia potencial receptor–ligante foram calculadas: *i*) aquela vinda exclusivamente de interações proteína–proteína, chamada de energia potencial intraproteína; *ii*) contribuição vinda de interações proteína–Ca^{2+}; e *iii*) contribuição de interações proteína–Na^{+}. A soma dessas três contribuições é a energia potencial receptor–

cátion. Uma vez que a sequência do domínio CTLD de cada um dos receptores tem tamanhos diferentes, os valores de energia potencial foram divididos pelo número de resíduos em cada caso para possibilitar comparações entre as três proteínas.

Na ausência de Ca^{2+}, CLEC5A tem os valores de energia potencial intraproteína mais negativos. A contribuição média de cada um de seus resíduos é de -17.16 ± 0.07 $kcal \cdot mol^{-1} \cdot res^{-1}$, aproximadamente 6.5 $kcal \cdot mol^{-1} \cdot res^{-1}$ mais negativo do que o DC-SIGN e MR (Fig. 14, porção preta das barras). A complexação de Ca^{2+} ao sítio principal do DC-SIGN e MR desestabiliza ligeiramente sua energia potencial intraproteína. No caso do DC-SIGN esse efeito aumenta progressivamente com a quantidade de Ca^{2+} complexados, chegando a um aumento de 2.5 $kcal \cdot mol^{-1} \cdot res^{-1}$ na energia intraproteína (Fig. 14, porção preta das barras). Isso se deve ao fato de que a complexação de Ca^{2+} aproxima cadeias laterais negativamente carregadas, gerando repulsão eletrostática. No entanto, compensando esse evento desestabilizador, as interações proteína–Ca^{2+} representam uma contribuição expressiva para a energia receptor–cátion: cada Ca^{2+} complexado contribui para a energia do complexo com aproximadamente -5 $kcal \cdot mol^{-1} \cdot res^{-1}$, o que corresponde a 50% da contribuição intraproteína. É interessante notar que quando DC-SIGN complexa três Ca^{2+} a maior parte da estabilização entálpica do complexo vem de interações com Ca^{2+} (Fig. 14, porção cinza das barras).

Verificamos que, na ausência de Ca^{2+} complexado aos receptores, Na^{+} da solução prontamente interage com os sítios principal e auxiliar. Diferente do Ca^{2+}, que sempre permanece complexado ao receptor quando está presente nas simulações, há múltiplos eventos de troca entre os Na^{+} complexados ao CTLD e os da solução, caracterizando uma interação transiente. Mesmo assim, sua contribuição para a energia do complexo é apreciável e estabiliza a energia proteína–cátion em aproximadamente 3 $kcal \cdot mol^{-1} \cdot res^{-1}$. Apesar de o sítio auxiliar do DC-SIGN complexar dois Ca^{2+}, e portanto ter mais resíduos disponíveis para ligação a cátions do que o sítio principal, sua interação com Na^{+} revela valores com significativa incerteza (Fig.14, parte hachurada e barras de erro correspondentes).

Esses resultados mostram uma relação interessante entre os receptores e os cátions. Na ausência de Ca^{2+}, CLEC5A tem a energia mais negativa entre os três e podemos considerá-lo como sendo o mais estável energeticamente —isso é, mais estável considerando apenas sua energia potencial— mesmo quando consideramos o efeito estabilizador da ligação de Ca^{2+} ao sítio principal do DC-SIGN e MR. Isso sugere que a incapacidade do CLEC5A ligar Ca^{2+} é compensada por interações intraproteína que diminuem sua energia intraproteína em relação ao DC-SIGN e MR. No entanto, a complexação de Ca^{2+} no sítio auxiliar do DC-SIGN supera a capacidade de estabilização interna do CLEC5A. Isso está de acordo com a visão atual de que o sítio principal tem função de complexação de carboidratos enquanto o auxiliar tem função de estabilizar os *loops* do mesmo sítio[73-75].

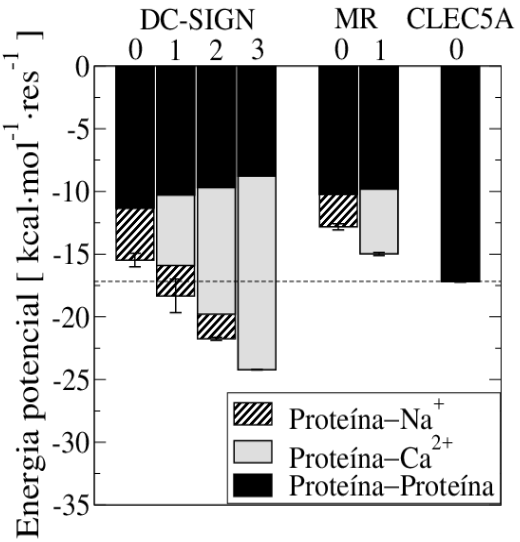

Figura 14: Energia potencial receptor–cátion dos complexos envolvendo os receptores DC-SIGN, MR e CLEC5A com diferentes quantidades de Ca^{2+} complexados a eles, conforme indicado no topo do gráfico: 0: nenhum Ca^{2+} complexado ao CTLD; 1: um Ca^{2+} complexado ao sítio principal; 2: dois Ca^{2+} ambos complexados ao sítio auxiliar; 3: três Ca^{2+} complexados ao CTLD, sendo um ao sítio principal e dois ao sítio auxiliar simultaneamente. Enquanto o tamanho total das barras representa a energia potencial do complexo receptor–cátions, suas subdivisões representam as contribuições da energia potencial intraproteína (área preta), interações proteína–Ca^{2+} (área cinza) e interações proteína–Na^{+} (área hachurada).

A integridade estrutural do sítio de reconhecimento de carboidratos depende de ligação de Ca^{2+} aos sítios principal e auxiliar no DC-SIGN e MR.

Após considerações energéticas sobre a interação CTL–Ca^{2+}, investigamos agora aspectos entrópicos através da verificação de como o Ca^{2+} modula a dinâmica estrutural dos receptores. A flexibilidade total das proteínas foi calculada através da soma da flutuação média quadrática (RMSF, do inglês "*Root Mean Square Fluctuation*") das posições dos carbonos-α de todos os resíduos. Foram considerados os seguintes estados de ligação a Ca^{2+}: nenhum Ca^{2+} complexado ao CTLD; um Ca^{2+} complexado ao sítio principal; dois Ca^{2+} ambos complexados ao sítio auxiliar; três Ca^{2+} complexados ao CTLD, sendo um ao sítio principal e dois ao sítio auxiliar simultaneamente.

Na ausência de Ca^{2+}, CLEC5A é o receptor mais rígido. Sua flexibilidade é aproximadamente 20% menor do que a do DC-SIGN e MR (Fig. 15, barras brancas). Enquanto a complexação de Ca^{2+} ao sítio principal do DC-SIGN e MR não causa mudança estatisticamente significante da flexibilidade dos receptores (Fig. 15, barras hachuradas), a ligação de todos os três Ca^{2+} ao DC-SIGN resulta em redução de 16% da sua flexibilidade, tornando-o quase tão rígido quanto o CLEC5A (Fig. 15, barras pretas). Ao contrário do

observado para a energia potencial, ligação de Ca^{2+} apenas ao sítio auxiliar do DC-SIGN não afeta significativamente sua flexibilidade (Fig. 15, barras cinzas).

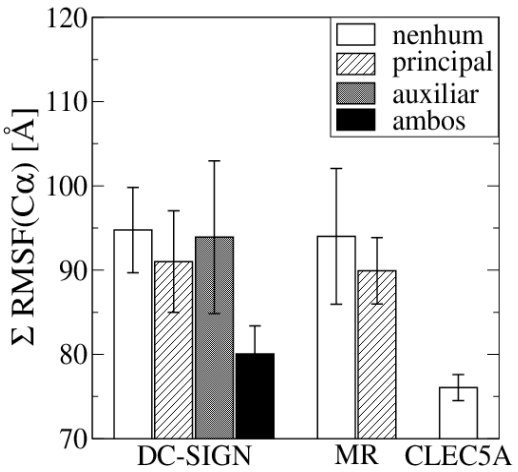

Figura 15: Flexibilidade total dos receptores conforme descrita pela somatória do RMSF dos carbonos-α de todos os seus resíduos. O estado de ocupação dos sítios de complexação de Ca^{2+} como descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas).

A fim de verificar quais são os resíduos mais afetados pela complexação de três Ca^{2+} ao DC-SIGN, calculamos a diferença de flexibilidade de cada resíduo, $\Delta RMSF_{resíduo}$, entre essa condição e o estado livre de Ca^{2+}. O efeito de redução de flexibilidade é totalmente local; apenas os *loops* que contém os sítios de complexação de Ca^{2+} são afetados (Fig. 16, painel superior). A alta incerteza na diferença de flexibilidade de alguns resíduos se deve a variações estruturais no estado livre de Ca^{2+}. Para o caso do MR é possível ver que o efeito da complexação de Ca^{2+} ao sítio principal também é localizado, embora as incertezas nos valores não permitam definir a extensão quantitativa do efeito (Fig. 16, painel inferior).

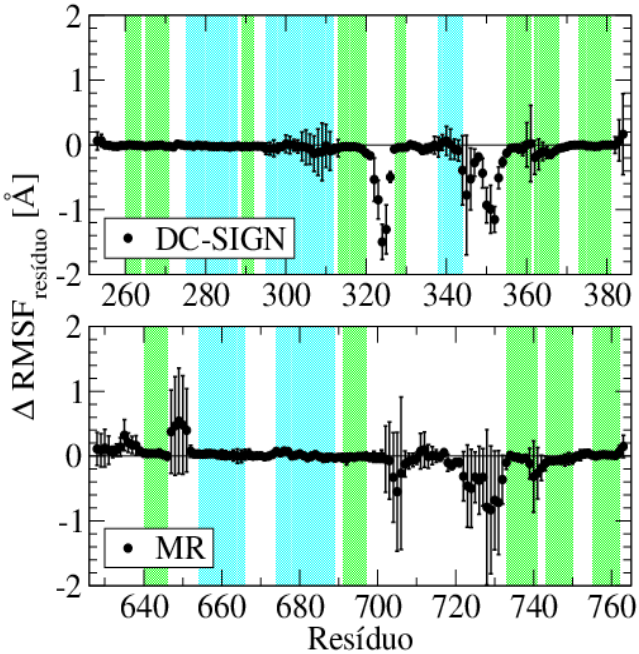

Figura 16: Diferença de flexibilidade de cada resíduo, $\Delta RMSF_{resíduo}$, após complexação de Ca^{2+}. Valores negativos indicam que o resíduo é menos flexível no estado saturado de Ca^{2+} do que no estado livre de Ca^{2+}. Áreas coloridas correspondem a segmentos de estrutura secundária com α–hélices indicadas em ciano e folhas–β em verde. **Painel superior:** Diferença de flexibilidade entre DC-SIGN complexando três e zero Ca^{2+}. **Painel inferior:** Diferença de flexibilidade entre MR complexando um e zero Ca^{2+}. As altas incertezas, indicadas pelas barras de erro, são consequência de variações estruturais no estado livre de Ca^{2+}.

Considerando esse resultado, a integridade estrutural do sítio de complexação de Ca^{2+} principal do DC-SIGN e MR, que está envolvido no reconhecimento de carboidratos, foi analisada através do comportamento dinâmico dos cinco resíduos que o compõe. Especificamente, os resíduos são Glu-347, Asn-349, Glu-354, Asn-365 e Asp-366 no DC-SIGN e Glu-725, Asn-727, Glu-733, Asn-747 e Asp-748 no MR. Para tal, examinamos o grau de alteração conformacional em relação à estrutura cristalográfica de referência e sua flexibilidade considerando todos os átomos pesados, ou seja, desvio médio quadrático (RMSD, do inglês "*Root Mean Square Deviation*") e soma do RMSF dos átomos pesados de tais resíduos.

Tanto o RMSF quanto o RMSD do sítio principal diminuem após a ligação de Ca^{2+} a ele, indicando diminuição de sua flexibilidade e que há algum grau de mudança conformacional em relação à estrutura de referência (Fig. 17). No caso do DC-SIGN, a ligação de Ca^{2+} somente a seu sítio auxiliar mostra efeito contrário ao do sítio principal; há aumento da flexibilidade e de sua distância conformacional em relação à estrutura cristalográfica, apesar do erro estatístico elevado devido à variações estruturais no estado livre de Ca^{2+}. (Fig. 17, barras cinzas). Já no caso da ligação de Ca^{2+} a ambos os sítios, principal e

auxiliar, ocorre um efeito não–aditivo; há grande diminuição de flexibilidade e RMSD (Fig. 17, barras pretas).

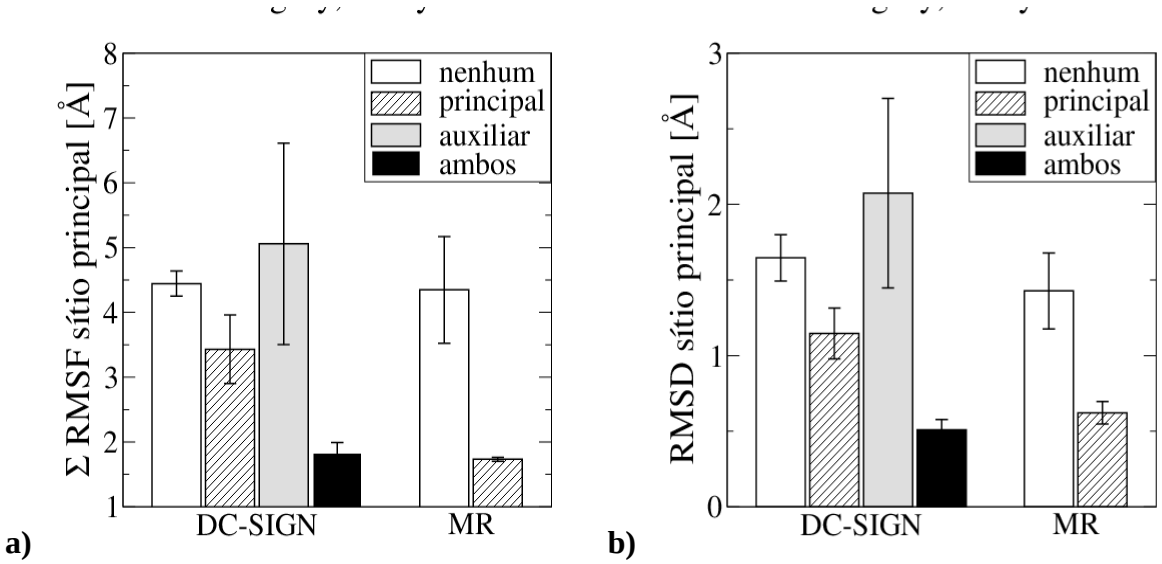

Figura 17: Integridade estrutural do sítio principal do DC-SIGN e MR conforme descrita pela **(a)** somatória dos RMSFs dos átomos pesados dos resíduos que os compõem; e **(b)** RMSD dos átomos pesados dos resíduos que os compõem em relação à estrutura cristalográfica de referência. O estado de ocupação dos sítios de complexação de Ca^{2+} está descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas).

Quando o comportamento dinâmico e conformacional dos cinco resíduos que compõem o sítio de complexação de Ca^{2+} principal são analisados separadamente, vemos que todos eles são modulados da mesma forma. Para DC-SIGN e MR, a ligação de Ca^{2+} apenas ao sítio principal atua como estabilizador estrutural do sítio de reconhecimento de açúcar (Fig. 18, barras hachuradas). Por outro lado, a ligação de Ca^{2+} exclusivamente ao sítio auxiliar do DC-SIGN tem efeito desestabilizador em quase todos os resíduos (Fig. 18, barras cinzas). Além disso, o sítio auxiliar tem efeito sinérgico com o principal: quando ambos complexam Ca^{2+} simultaneamente há um aumento significativo do efeito estabilizador observado para o principal (Fig. 18, barras pretas). Portanto, nossos resultados indicam que a manutenção do sítio de reconhecimento de açúcar em sua conformação ativa —isso é, que está apta a estabelecer interações específicas com carboidratos— depende da ligação de Ca^{2+} ao sítio principal e, nas CTLD em que existir, ao auxiliar simultaneamente.

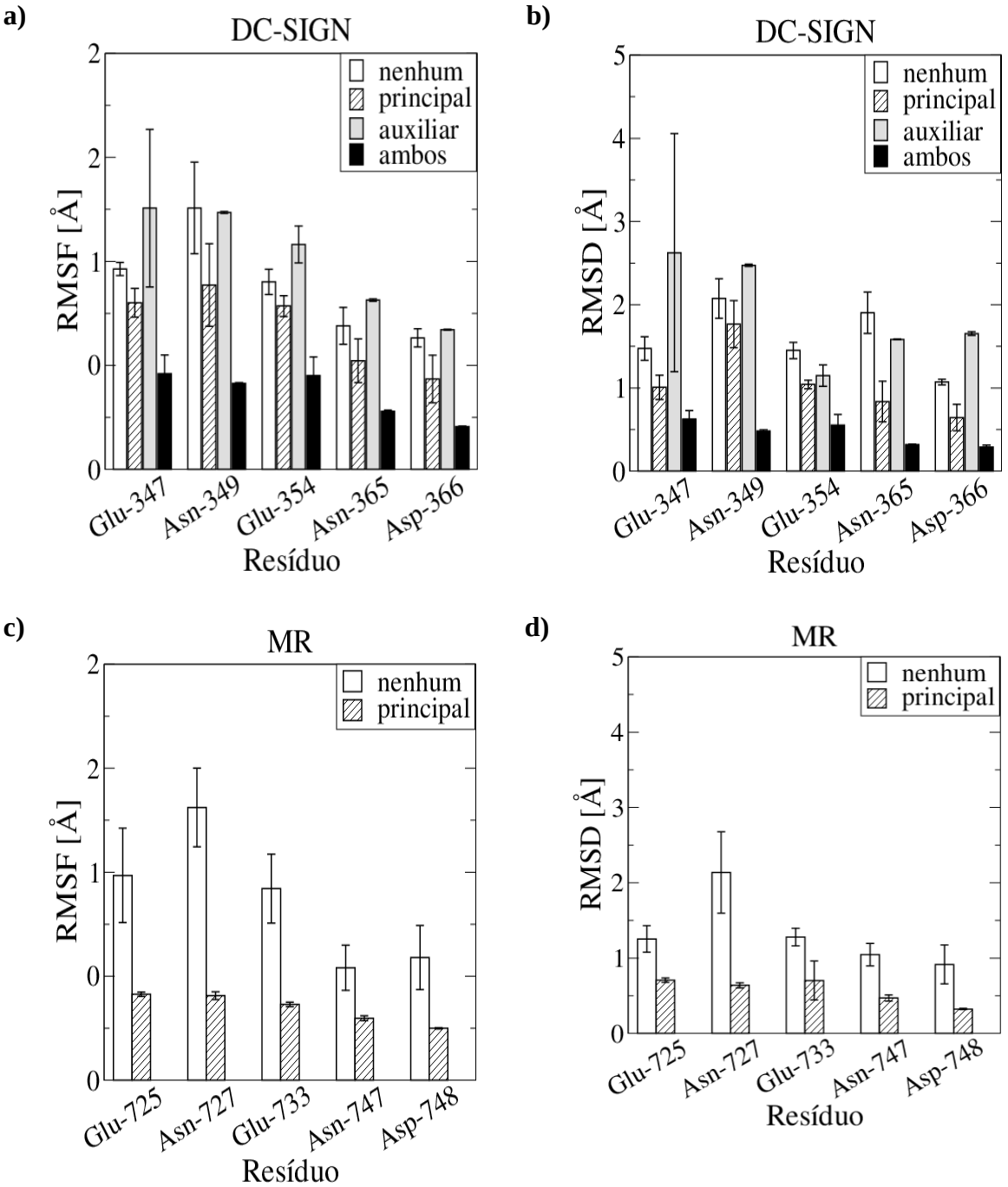

Figura 18: Integridade estrutural do sítio principal do **(a,b)** DC-SIGN e **(c,d)** MR conforme descrita pela **(a,c)** RMSF dos átomos pesados dos resíduos que os compõem; e **(b,d)** RMSD dos átomos pesados dos resíduos que os compõem em relação à estrutura cristalográfica de referência. O estado de ocupação dos sítios de complexação de Ca^{2+} conforme descrito na legenda: nenhum sítio ocupado (barras brancas), apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios ocupados (barras pretas).

Na ausência de Ca^{2+}, o sítio principal do DC-SIGN e do MR adota uma conformação aberta na qual consegue interagir com até três Na^{+} da solução ao mesmo tempo (Fig. 19). Vale ressaltar que há trocas contínuas entre os Na^{+} que estão complexados ao CTLD

e aqueles livres na solução, caracterizando a natureza de vida curta da interação receptor–Na^+, o que está de acordo com o observado para as altas incertezas na contribuição dessa interação para a energia potencial intracomplexo (Fig. 19).

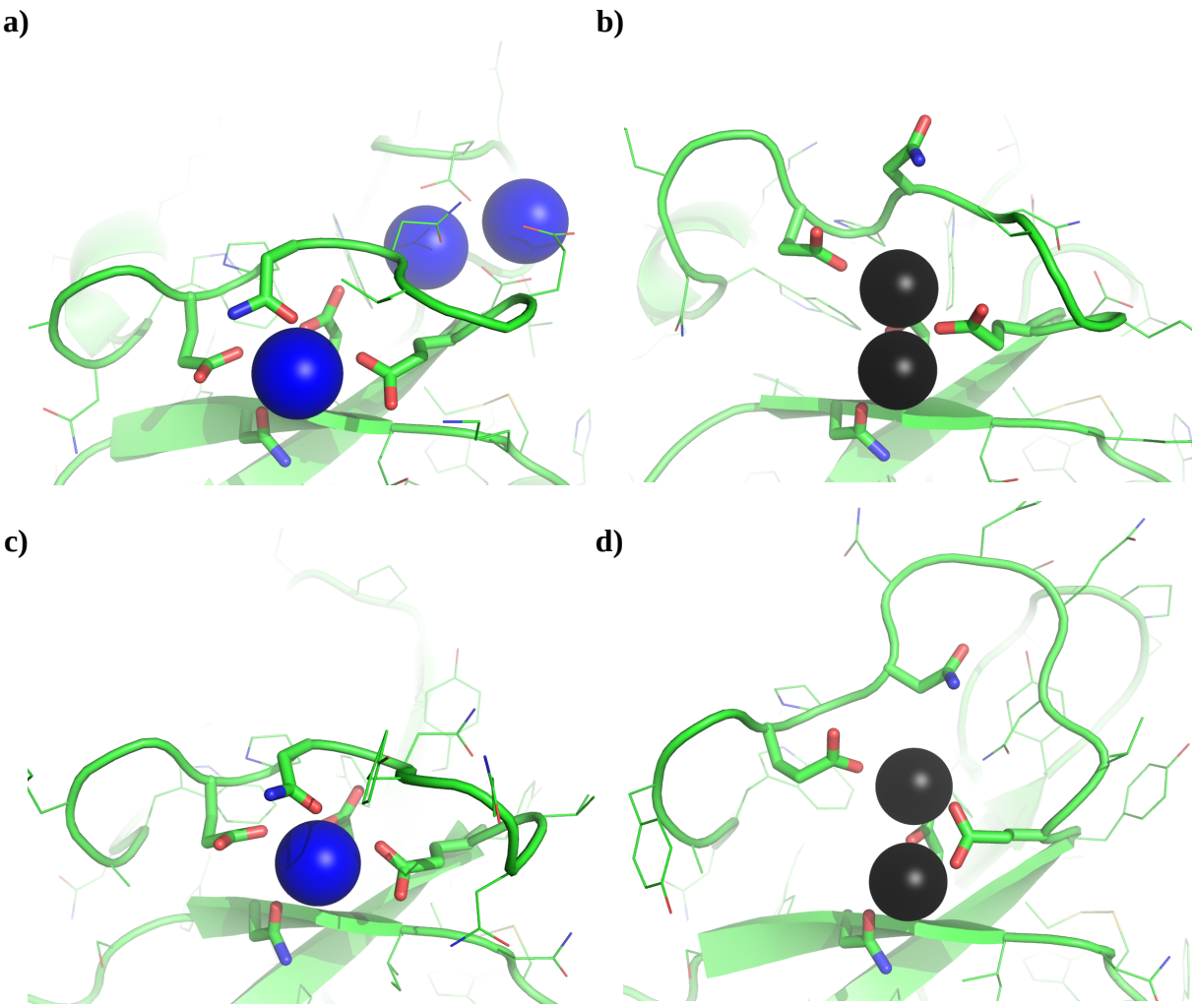

Figura 19: Visualização da integridade estrutural do sítio principal do **(a,b)** DC-SIGN e **(c,d)** MR no estado saturado de Ca^{2+} e **(a,c)** na ausência de Ca^{2+}. *Snapshots* obtidos após mais de 150 ns de simulação de dinâmica molecular. A proteína está mostrada em *cartoon* verde com as cadeias laterais dos resíduos do sítio principal destacadas como *sticks* e as demais mostradas como linha. Ca^{2+} é mostrado como esferas azuis e Na^+ como esperas cinza escuro.

Trabalhos da literatura tem proposto que CRD-4 do MR sofre mudança conformacional ao complexar Ca^{2+} com base na observação de que a presença do cátion torna o domínio resistente à ação de proteases[85-88]. Adicionalmente, foi demonstrado por cristalografia de raios-X que o Ca^{2+} pode mudar a conformação e flexibilidade do sítio de reconhecimento de açúcar do CD23, uma CTL que pertence ao mesmo grupo dentro da superfamília CTLD que o DC-SIGN[159]. Nossos resultados fornecem, através de uma abordagem molecular detalhada, novas evidências de que a complexação de Ca^{2+} pode modular tanto a conformação quanto a dinâmica das CTLs dos grupos II e VI.

Apesar de não se saber onde o sítio de interação com ligantes do CLEC5A está localizado, o fato de que seu perfil de flexibilidade é quase idêntico ao do DC-SIGN quando complexado a três Ca^{2+} (Fig. 20) sugere, novamente, que a falta de capacidade do CLEC5A ligar Ca^{2+} é compensada por interações intraproteína que o tornam menos flexível que os outros receptores considerados. Tal compensação pode ser uma alternativa constitutiva para a complexação de Ca^{2+} como um mecanismo de redução da entropia conformacional para facilitar a interação com seus ligantes.

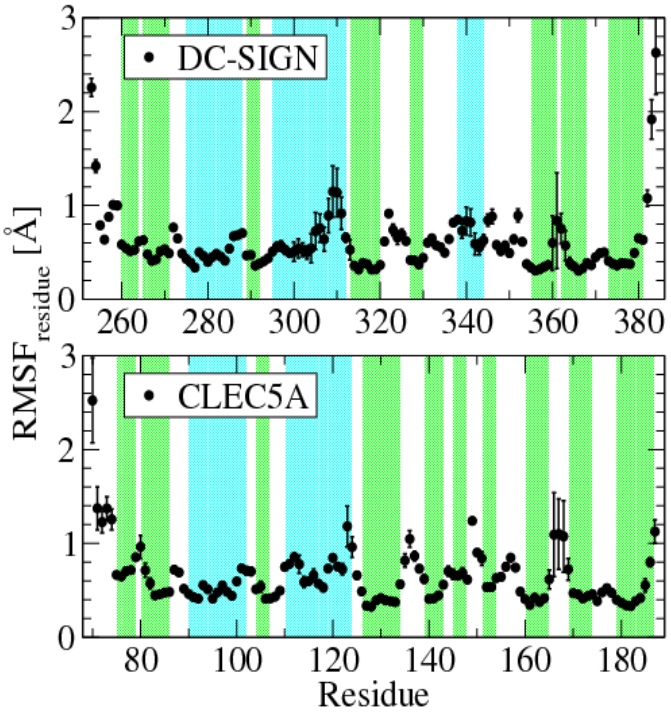

Figura 20: Perfil de flexibilidade por resíduo, $RMSF_{resíduo}$. Áreas coloridas correspondem a segmentos de estrutura secundária com α–hélices indicadas em ciano e folhas–β em verde. **Painel superior:** DC-SIGN quando complexado a três Ca^{2+}. **Painel inferior:** CLEC5A.

Os sítios de complexação de Ca^{2+} do DC-SIGN são independentes

A observação de que a ligação de Ca^{2+} a ambos os sítios do DC-SIGN tem efeito sinérgico sobre a diminuição local da flexibilidade do receptor levanta a possibilidade de que a complexação de Ca^{2+} seja cooperativa entre eles. Isso foi investigado através do cálculo da força de ruptura de cada cátion em diferentes estados de ligação a Ca^{2+}.

Não há influência significativa da ocupância de um sítio sobre a afinidade relativa do outro por Ca^{2+} (Fig. 21), sugerindo que eles são independentes. Foi possível estabelecer a escala relativa de afinidade de cada um por Ca^{2+}: O principal é o que tem maior afinidade, seguido pelo cátion mais interno do sítio auxiliar, com força de ruptura 10% menor,

e então pelo cátion mais externo do sítio auxiliar, com força de ruptura 35% menor do que o principal. Tal escala de afinidade sugere que essa seria a ordem de ocupação dos sítios no sistema biológico. Adicionalmente, a afinidade do sítio principal do MR por Ca^{2+} é a mesma que a do DC-SIGN (Fig. 21).

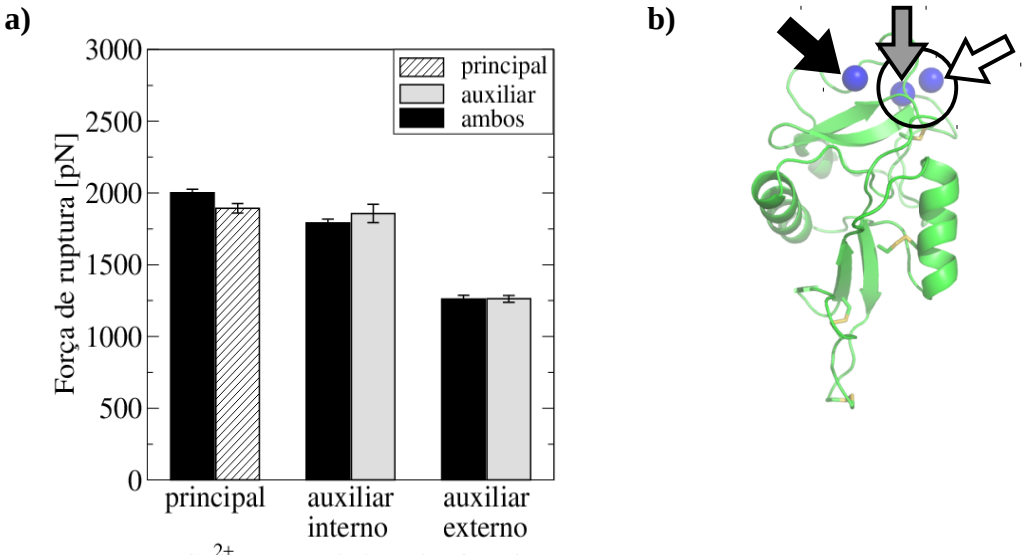

Figura 21: a) Força de ruptura para remoção de cada Ca^{2+} do DC-SIGN, conforme especificado no eixo x. O estado de ocupação dos sítios de complexação de Ca^{2+} está descrito na legenda: apenas sítio principal ocupado (barras hachuradas), apenas sítio auxiliar ocupado (barras cinzas) e ambos os sítios principal e auxiliar ocupados (barras pretas). **b)** Estrutura de referência do DC-SIGN. O sítio principal está indicado por uma seta preta e o auxiliar por um círculo preto. Os Ca^{2+} que fazem parte do sítio auxiliar são individualmente chamados de auxiliar interno e externo, indicados por uma seta cinza e branca, respectivamente.

A interação direta entre Ca^{2+} e carboidratos opõe-se à formação do complexo

É aceito na literatura que o mecanismo molecular da atividade lectínica Ca^{2+}–dependente das CTLs envolve a coordenação do Ca^{2+} principal por hidroxilas vizinhas no anel do açúcar. Nossos resultados destacam outro aspecto da relação Ca^{2+}–açúcar: a complexação de Ca^{2+} é essencial para estabilizar os resíduos do sítio de reconhecimento de carboidratos na sua conformação ativa. Portanto, a fim de verificar se o papel majoritário do Ca^{2+} nas CTLs é interagir diretamente com o carboidrato ou estabilizar o sítio principal, analisamos a extensão na qual a interação Ca^{2+}–açúcar contribui diretamente para a afinidade por carboidratos.

Nossa abordagem foi usar o método MM/PBSA para calcular a contribuição do Ca^{2+} principal para a energia livre de interação entre DC-SIGN e dois tetrassacarídeos representativos de seus ligantes naturais: a ramificação externa α1,6 da glicosilação *high-mannose* sem as manoses α1,2 terminais (Man_4) e o antígeno Lewis-x (Le^x), que são representativos dos padrões de glicosilação *high-mannose* associado a patógenos e da glicosilação *complex-type* associada a antígenos próprios, respectivamente. Adicionalmente,

como um argumento comparativo, fizemos os mesmos cálculos considerando a substituição do Ca^{2+} por Na^{+}. O cátion Na^{+} foi escolhido para representar a ausência de Ca^{2+} porque é o mais abundante no meio extracelular e porque verificamos que Na^{+} da solução prontamente interage com o sítio de complexação de Ca^{2+} na ausência deste.

O estado de referência foi obtido para ambos os açúcares extraindo-se 400 *frames* de trajetórias obtidas por simulação de dinâmica molecular do DC-SIGN em complexo com os tetrassacarídeos mencionados. Um novo estado foi criado pela substituição direta do Ca^{2+} complexado no sítio principal por Na^{+} nesses mesmos 400 *frames*. Portanto, para cada complexo receptor–carboidrato foram obtidas duas trajetórias idênticas exceto pela identidade do cátion presente no sítio principal. Essas trajetórias representam a situação hipotética na qual a conformação do sítio principal seria a mesma quando DC-SIGN complexa Ca^{2+} ou Na^{+}. Apesar da artificialidade da condição na qual há Na^{+} no sítio principal, consideramos que, para fins comparativos, trata-se de uma solução melhor do que deixar o sítio sem cátion algum. Assim, caso a interação direta entre o Ca^{2+} principal e os carboidratos seja determinante para a afinidade CTL–açúcar, sua contribuição para $\Delta G_{\text{interação}}$, definida aqui como $\Delta G(\text{cátion} \mid \text{Ca}^{2+})$, deverá ser consideravelmente negativa. Além disso, a diferença na contribuição do cátion devido à mudança de Ca^{2+} para Na^{+}, definida como $\Delta\Delta G(\text{cátion}) = \Delta G(\text{cátion} \mid \text{Na}^{+}) - \Delta G(\text{cátion} \mid \text{Ca}^{2+})$, deverá ser positiva porque se espera que $\Delta G(\text{cátion} \mid \text{Na}^{+})$ seja menos negativo do que $\Delta G(\text{cátion} \mid \text{Ca}^{2+})$. Por outro lado, se o papel majoritário do Ca^{2+} no sítio principal for estabilizar seus resíduos na conformação ativa, $\Delta G(\text{cátion} \mid \text{Ca}^{2+})$ e $\Delta\Delta G(\text{cátion})$ deverão ser próximos de zero.

Curiosamente, observamos que $\Delta\Delta G(\text{cátion})$ é negativo para ambos os carboidratos, sendo -23.8 ± 0.3 kcal·mol^{-1} para Man_4 e -30.2 ± 0.2 kcal·mol^{-1} para Le^X, indicando que a interação direta Na^{+}–açúcar seria preferível à Ca^{2+}–açúcar. Esse é um resultado inesperado uma vez que é sabido que Ca^{2+} é um fator importante para a afinidade de carboidratos pelo DC-SIGN e outras CTL. Então, a fim de entender a origem de tal resultado intrigante, analisamos os componentes de $\Delta\Delta G(\text{cátion})$ de acordo com o método MM/PBSA, isso é: os termos de solvatação polar, solvatação apolar e mecânica molecular, que é a soma das energias devido a interações eletrostáticas e de dispersão.

Como esperado, a contribuição da mecânica molecular de ambos os cátions para ΔG_{Man_4} é favorável à formação do complexo e aquela correspondente ao Na^{+} é aproximadamente metade daquela do Ca^{2+}: -10.2 ± 0.1 e -19.6 ± 0.2 kcal·mol^{-1}, respectivamente. Isso é um reflexo do domínio eletrostático sobre o termo da mecânica molecular, que também inclui interações de dispersão. Resultados para Le^X são similares (Fig. 22, barras azuis).

A contribuição da solvatação apolar é pequena e similar, -0.04 kcal·mol^{-1}, para todos os casos. Por outro lado, a solvatação polar aparece como a causa do comportamento inesperado: durante a formação do complexo CTL–açúcar, o Ca^{2+} principal é removido do contato com o solvente, resultado na contribuição desfavorável da solvatação polar para $\Delta G(\text{cátion} \mid Ca^{2+})$ de 38.2 ± 0.3 kcal·mol^{-1}, enquanto a penalidade de dessolvatação para Na^{+} é de apenas 10% desse valor, 4.9 ± 0.2 kcal·mol^{-1}. O mesmo acontece para Le^{X} (Fig. 22, barras vermelhas).

Para Man_4, a contribuição total do Ca^{2+} para a energia livre de interação com DC-SIGN, $\Delta G(\text{cátion} \mid Ca^{2+})$, é de 18.6 ± 0.3 kcal·mol^{-1} e a contribuição total do Na^{+}, $\Delta G(\text{cátion} \mid Na^{+})$, é -5.3 ± 0.2 kcal·mol^{-1}, resultando no valor negativo de $\Delta\Delta G(\text{cátion})$ descrito anteriormente. Resultados para Le^{X} seguem a mesma tendência. Ainda mais interessante do que o valor negativo de $\Delta\Delta G(\text{cátion})$ é que $\Delta G(\text{cátion} \mid Ca^{2+})$ é positivo enquanto $\Delta G(\text{cátion} \mid Na^{+})$ é negativo, indicando que a interação direta Ca^{2+}–carboidrato se opõe à formação do complexo CTL–açúcar (Fig. 22, barras pretas), o que vai totalmente contra a visão atual da literatura.

A fim de verificar a eventual presença de artefatos devido ao método, repetimos os cálculos de MM/PBSA variando o valor da constante dielétrica interna da proteína de seu valor-padrão, 1, até 8. Os resultados são robustos e indicam que a contribuição da mecânica molecular torna-se desprezível conforme valores de constante dielétrica interna maiores são usados e que a solvatação polar acaba por dominar o valor de $\Delta\Delta G(\text{cátion})$ (Fig. 22). Essa observação reforça nosso achado de que a penalidade de dessolvatação do Ca^{2+} principal faz com que sua interação direta com açúcares se oponha à formação do complexo.

Finalmente, para verificar se o mesmo pode ser replicado em outras CTLs, fizemos as mesmas simulações de dinâmica molecular e cálculos de MM/PBSA para o complexo MR–Man_4 construído a partir de sua estrutura cristalográfica sem ligante. Todos os resultados são muito similares ao obtido para o DC-SIGN (Fig. 22b, 22d, 22f) e todos os achados são igualmente aplicáveis ao complexo MR–Man_4, sugerindo que nossos resultados refletem um aspecto fundamental da interação entre membros da família CTL e açúcares.

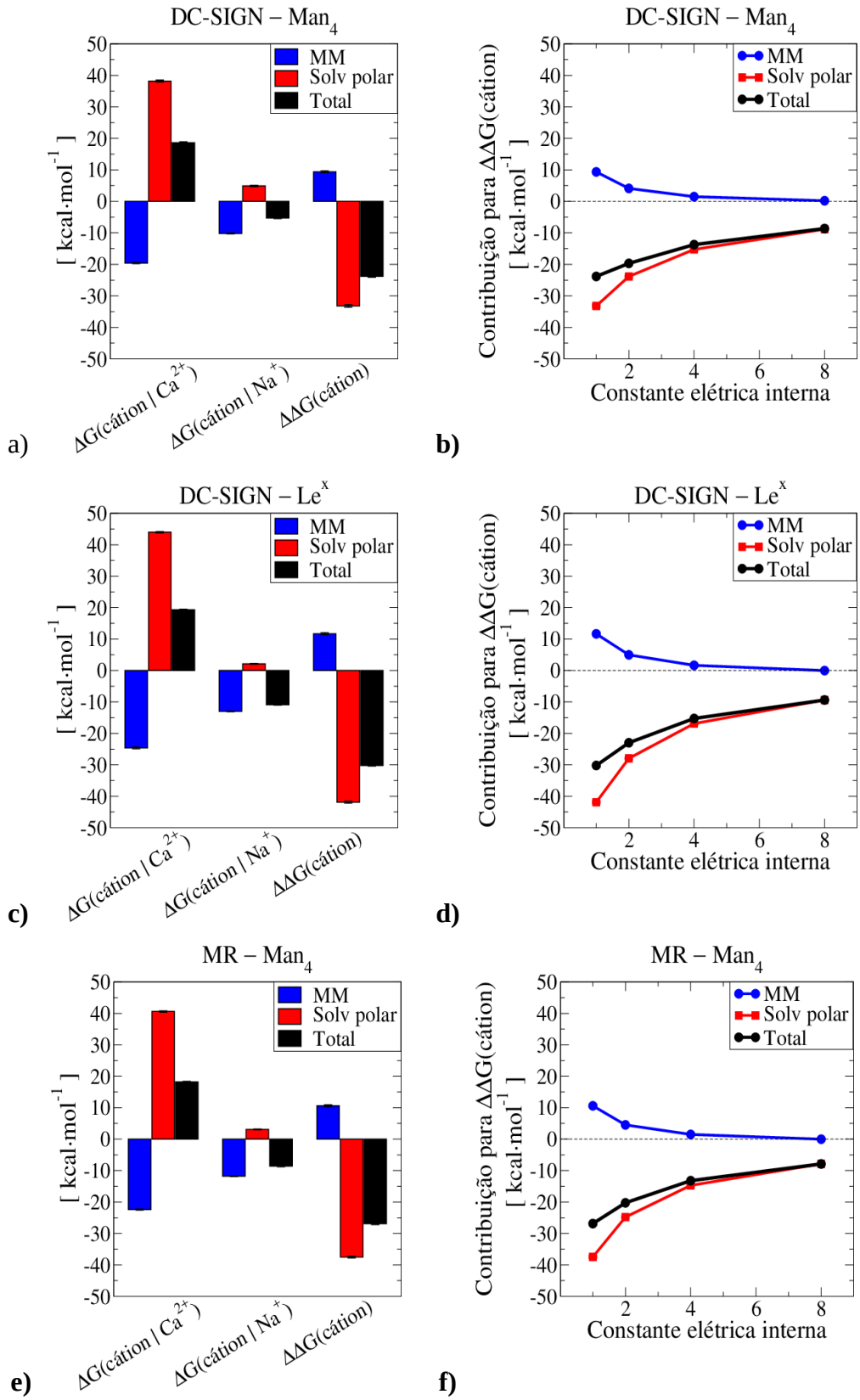

Figura 22: (a,c,e) Contribuições para $\Delta G(\text{cátion})$ e $\Delta\Delta G(\text{cátion})$: mecânica molecualar (azul), solvatação polar (vermelho) e total (preto). O valor usado para a constante dielétrica interna da proteína foi 1. **(b,d,f)** $\Delta\Delta G(\text{cátion})$ em função da constante dielétrica interna do receptor.

A presença de Ca^{2+} modula a afinidade dos resíduos do sítio principal por carboidratos

Conforme descrito na seção anterior, a diferença da contribuição direta do cátion para a energia livre de interação com açúcares, $\Delta\Delta G(\text{cátion})$, favorece fortemente Na^+ sobre Ca^{2+}. Mesmo assim, a diferença total —isso é, considerando tanto as contribuições dos resíduos do receptor e do cátion complexado ao sítio principal— da energia livre de interação entre as trajetórias na qual há Na^+ ou Ca^{2+} complexado no sítio principal, $\Delta\Delta G(\text{total}) = \Delta G(\text{total}\,|\text{Na}^+) - \Delta G(\text{total}\,|\text{Ca}^{2+})$, é pequena: -2.5 ± 0.6 kcal·mol^{-1} e 2.0 ± 0.6 kcal·mol^{-1} para DC-SIGN interagindo com Man_4 e Le^X, respectivamente, e 0 ± 1 kcal·mol^{-1} para MR interagindo com Man_4.

Isso acontece porque a contribuição dos resíduos que compõem o sítio principal para a energia livre de interação, $\Delta G(\text{resíduo})$, é modulada pela presença de Ca^{2+}. As propriedades físico-químicas do sítio de reconhecimento de açúcar mudam de acordo com o cátion presente no sítio principal de forma que a contribuição energética dos cinco resíduos que o compõem favorece fortemente Ca^{2+} em vez de Na^+, isso é, $\Delta\Delta G(\text{resíduos})$ é positivo (Fig. 23, barras pretas). Portanto, $\Delta\Delta G(\text{resíduos})$ opõe-se ao $\Delta\Delta G(\text{cátion})$ desfavorável, levando a diferenças na energia livre total, $\Delta\Delta G(\text{total}) = \Delta\Delta G(\text{cátion}) + \Delta\Delta G(\text{resíduos})$, pequenas ou nulas.

Em oposição direta ao que observamos para contribuições do cátion para a energia livre, as contribuições dos resíduos do sítio principal são favoráveis na presença de Ca^{2+} (Fig. 23, barras cinzas) e, exceto para Glu-347, são desfavoráveis na presença de Na^+ (Fig. 23, barras hachuradas). Portanto, apesar de o Ca^{2+} principal opor-se diretamente à complexação de carboidratos devido a seu elevado custo de dessolvatação durante a formação do complexo, ele favorece a interação dos resíduos do sítio principal com carboidratos através da modulação das características físico-químicas do sítio.

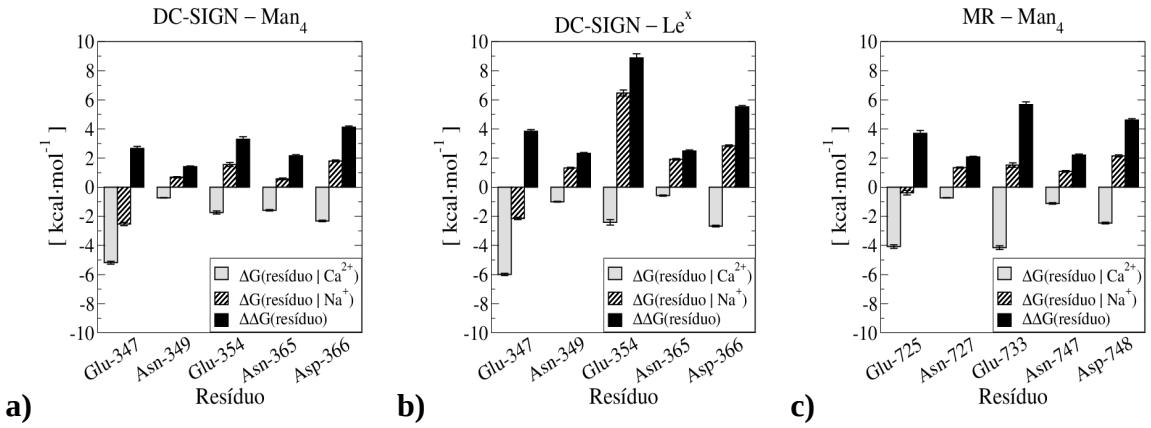

Figura 23: Contribuições dos resíduos que compõem o sítio principal para a energia livre de interação CTL–carboidrato quando há Ca^{2+} (barras cinzas) ou Na^{+} no sítio principal (barras hachuradas) e a diferença entre tais contribuições (barras pretas). **(a)** DC-SIGN – Man_4; **(b)** DC-SIGN – Le^X; **(c)** MR – Man_4.

Pelo menos duas moléculas de água são removidas da primeira camada de solvatação do Ca^{2+} principal para a formação do complexo CTL–carboidrato

A fim de obter mais detalhes sobre o mecanismo molecular do custo de solvatação imposto pelo Ca^{2+} principal durante a formação do complexo CTL–carboidrato, trajetórias obtidas por simulação de dinâmica molecular do DC-SIGN e MR complexando Ca^{2+} foram usadas para calcular a densidade local de água ao redor da proteína. Após remoção dos movimentos de translação e rotação do receptor, o espaço foi mapeado por uma grade cúbica com resolução de 1 $Å^3$ e o número médio de moléculas de água dentro de cada célula da grade durante a simulação foi calculado. Sítios específicos podem ser identificados comparando seu valor de densidade local com a densidade média da caixa inteira e seu desvio padrão.

Tanto para DC-SIGN quanto MR foram identificados sítios específicos onde a densidade local de água é significativamente alta, sendo é até 33 desvios padrão maior do que a densidade média da caixa toda (Fig. 24a, 24d). Em particular, há três sítios de alta densidade local de solvente ao redor do Ca^{2+} principal, indicando a localização das moléculas de água que fazem parte da primeira camada de solvatação do cátion. Além da interação direta com o Ca^{2+}, moléculas de água que estiverem nesses sítios conseguem estabelecer redes de ligação de hidrogênio com os resíduos do sítio principal. No caso do DC-SIGN, os sítios da primeira camada de solvatação do Ca^{2+} têm densidade local de água 11, 13 e 33 desvios padrão maior do que a média da caixa (Fig. 24b). Os resultados são relativamente similares para o MR: tais sítios têm densidade local 10, 13 e 17 desvios padrão maior do que a média (Fig. 24e).

Dois dos sítios da primeira camada de solvatação do Ca^{2+} principal correspondem a posições ocupadas pelas hidroxilas do anel do monossacarídeo que interage com o sítio principal. Essa correspondência, revelada através da sobreposição das estruturas

cristalográficas do DC-SIGN à análise de densidade local de água, sugere que a formação do complexo exige a remoção de pelo menos duas moléculas da primeira camada de solvatação do Ca^{2+} principal (Fig. 24 c,f).

De forma geral, o deslocamento de moléculas de água que inicialmente fazem parte das camadas de solvatação de um receptor durante a formação de um complexo pode tanto favorecer quanto desfavorecer o processo. Nos casos em que as moléculas de água não conseguem estabelecer ligações de hidrogênio efetivas com o soluto, como no caso de moléculas hidrofóbicas, a liberação dessas moléculas de água favorece a formação do complexo devido ao aumento da entropia translacional e rotacional das moléculas de água liberadas para o solvente[160,161]. No entanto, nos casos como o do Ca^{2+} principal em que há interações específicas entre as moléculas de água de solvatação e o soluto, o deslocamento dessas moléculas para o solvente desfavorece a formação do complexo porque o seu ganho de entropia não compensa a perda de entalpia devido às interações soluto–água rompidas [162,163].

Foi possível identificar o mecanismo molecular do alto custo de dessolvatação identificado pelo método MM/PBSA através de um método completamente independente, fortalecendo a nova visão que estamos propondo para as relações entre CTLs, Ca^{2+} e carboidratos. Adicionalmente, o cálculo da densidade local de água pode ser útil para encontrar locais de interesse para interações de candidatos a inibidor de tais receptores[164].

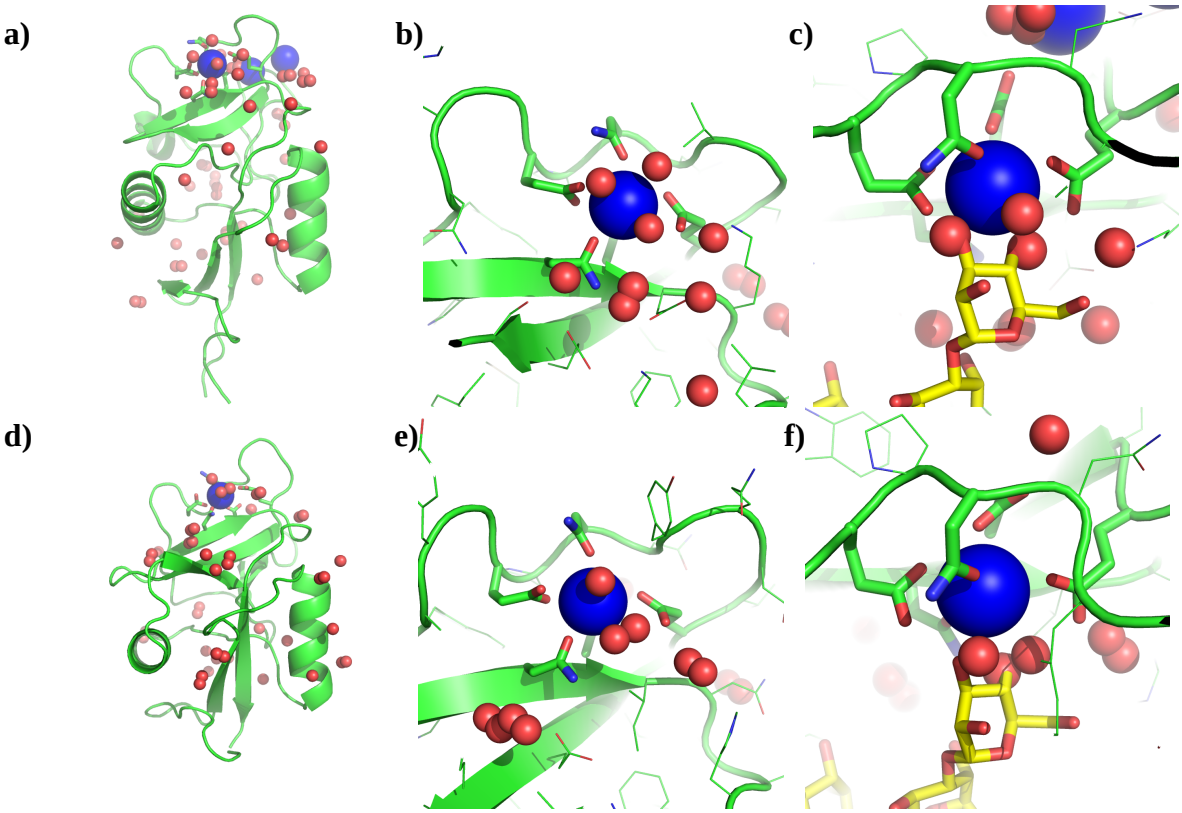

Figura 24: Densidade local de água ao redor do **(a, b, c)** DC-SIGN e **(d, e, f)** MR. **Primeira coluna:** sítios onde a densidade local de água é pelo menos 10 desvios padrão maior do que a média da caixa estão representados por pequenas esferas vermelhas. **Segunda coluna:** Sítios de alta densidade local de água próximos do Ca^{2+} principal, evidenciando a identificação da estrutura da sua primeira camada de solvatação. **Terceira coluna:** Sobreposição da estrutura cristalográfica do DC-SIGN em complexo com Man_4 à análise de densidade de água. As hidroxilas do anel do carboidrato que interage com o Ca^{2+} principal ocupam dois dos seus sítios de solvatação.

4.2. Mecanismos de reconhecimento CTL–*high-mannose*

Após a investigação da interação entre CTLs e Ca^{2+}, prosseguimos para a caracterização dos mecanismos moleculares do reconhecimento da glicosilação *high-mannose*, que está associada a patógenos, pelas CTLs DC-SIGN e MR. O CLEC5A não foi incluído nesta etapa porque há evidências de que ele não complexa carboidratos[100].

Ambos os receptores foram simulados em complexo com a porção da glicosilação *high-mannose* que contém apenas manoses. Especificamente, os fragmentos Man_5 e Man_9 foram usados para representar a menor e maior glicosilação *high-mannose*, que são $GlcNAc_2Man_5$ e $GlcNAc_2Man_9$, respectivamente (Fig. 25). Isso foi feito porque, no sistema biológico, os resíduos de GlcNAc estão muito próximos da proteína glicosilada —a glicoproteína do envelope do vírus da Dengue no nosso caso. Por isso, espera-se que esses açúcares não tenham interação efetiva com as lectinas.

Os complexos foram construídos a partir da estrutura cristalográfica do DC-SIGN em complexo com Man_4 (PDB ID 1SL4). As posições do açúcar foram usadas como referência para construção dos fragmentos Man_5 e Man_9. Apesar de não haver estrutura cristalográfica do MR em complexo com açúcares, as características bioquímicas de seu sítio de ligação a Ca^{2+} e estruturas de proteínas relacionadas a ele sugerem que sua interação com carboidratos é muito parecida com a observada para o DC-SIGN[73,90]. Assim, a estrutura cristalográfica do DC-SIGN em complexo com Man_4 (PDB ID 1SL4) foi sobreposta à estrutura do MR sem ligante (PDB ID 1EGG) e o mesmo procedimento foi seguido para construção de seus complexos com Man_5 e Man_9.

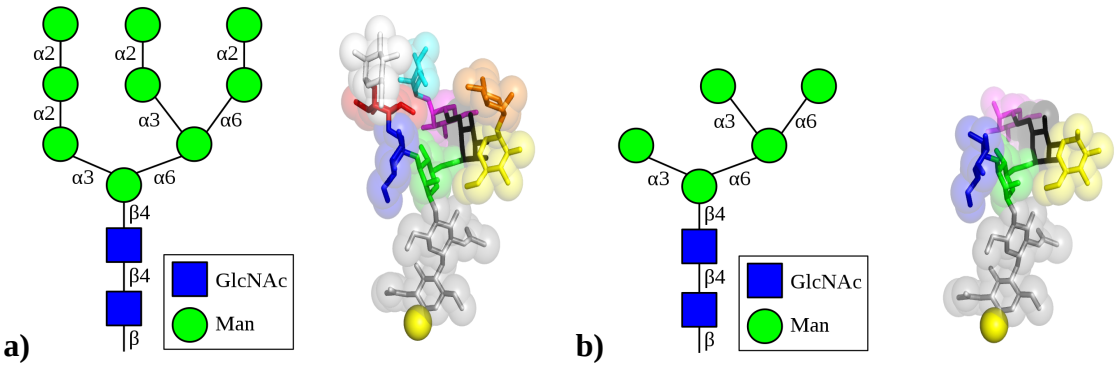

Figura 25: Estrutura da glicosilação *high-mannose*. Esquerda: Representação esquemática com os tipos de ligação glicosídica indicados; direita: representação atomística na mesma orientação usada na representação esquemática. Os resíduos de GlcNAc estão em cinza e cada manose está representada em uma cor diferente. A esfera amarela representa o ponto onde a glicosilação é covalentemente ligada à glicoproteína do envelope do vírus da Dengue. **a)** Glicosilação maior, $GlcNAc_2Man_9$. **b)** Glicosilação menor, $GlcNAc_2Man_5$.

4.2.1. DC-SIGN

Inicialmente buscamos interações hidrofóbicas ou de empilhamento de anéis que possam contribuir para a interação entre DC-SIGN e os fragmentos da glicosilação *high-mannose*. Isso foi feito através do cálculo da área enterrada de cada resíduo, definido aqui como a diferença entre sua área exposta ao solvente na ausência e na presença do carboidrato. O resíduo Phe-313 é o único hidrofóbico ou aromático que é significativamente enterrado devido à interação com carboidratos (Fig. 26). O valor de sua área enterrada, 62 ± 4 $Å^2$, é mais do que o dobro do que o próximo resíduo mais afetado, Ser-360. A importância de Phe-313 no reconhecimento de carboidratos pelo DC-SIGN está de acordo o observado através de cristalografia de raios-X[73].

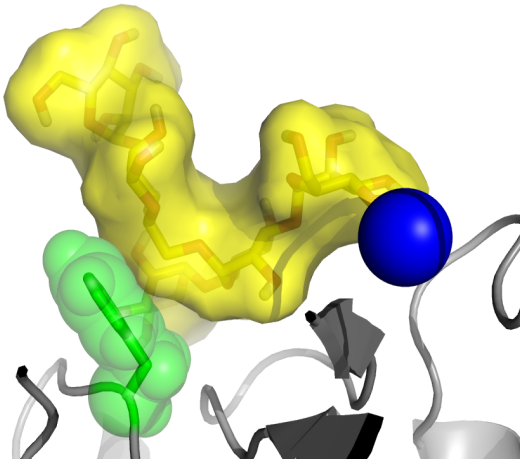

Figura 26: Interação entre o fragmento Man_5, representado como *sticks* e superfície amarela, com o resíduo Phe-313 do DC-SIGN, destacado em *sticks* e esferas verdes. A proteína está mostrada em *cartoon* cinza e o Ca^{2+} principal como uma esfera azul.

Outro tipo de interação que confere estabilidade e contribui para a especificidade do reconhecimento da glicosilação *high-mannose* pelo DC-SIGN é a ligação de hidrogênio. Foi observado que ao longo das simulações o DC-SIGN faz, em média, 6±1 ligações de hidrogênio com Man_5 e 7±1 com Man_9.

A decomposição do total de ligações de hidrogênio entre DC-SIGN e *high-mannose* na contribuição de cada um dos resíduos de manose separadamente, numerados de Man-1 até Man-9 (Fig. 27), revela que apenas dois deles interagem consistentemente com a lectina ao longo das simulações (Tabela 2): Man-6, que é a manose principal, e Man-8 (Fig 28). É interessante notar que Man-5 estabelece mais ligações de hidrogênio com DC-SIGN em Man_5 do que em Man_9 provavelmente devido a ajustes conformacionais necessários para acomodar os monossacarídeos extras presentes em Man_9. Essas manoses adicionais presentes em Man_9 tem dois tipos de comportamento: Man-3, -4 e -7 não interagem significativamente com o receptor enquanto Man-9 parece ser capaz de estabelecer interações transitórias (Tabela 2).

Tabela 2: Número médio de ligações de hidrogênio estabelecidas entre DC-SIGN e cada um dos resíduos de manose dos fragmentos Man_5 e Man_9 ao longo das simulações.

Man-x	Man_5	Man_9
1	0.1 ± 0.1	0.06 ± 0.01
2	0.4 ± 0.3	0.3 ± 0.3
3		0.1 ± 0.1
4		0.3 ± 0.3
5	1.2 ± 0.2	0.4 ± 0.3
6	2.4 ± 0.5	2.7 ± 0.5
7		0.004 ± 0.002
8	1.9 ± 0.2	2.3 ± 0.2
9		0.8 ± 0.6

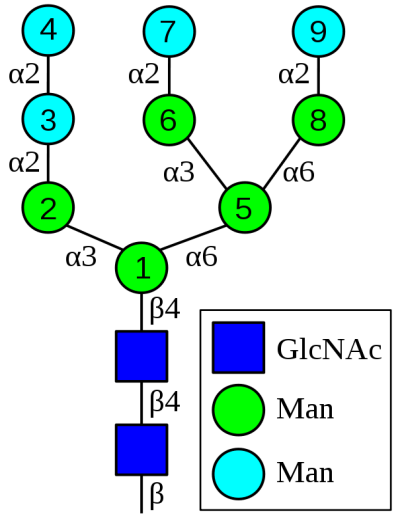

Figura 27: Estrutura esquemática da glicosilação $GlcNAc_2Man_9$ e a numeração de cada resíduo de manose que adotamos. $GlcNAc_2Man_5$ não tem as manoses 3, 4, 7 e 9, destacadas em ciano.

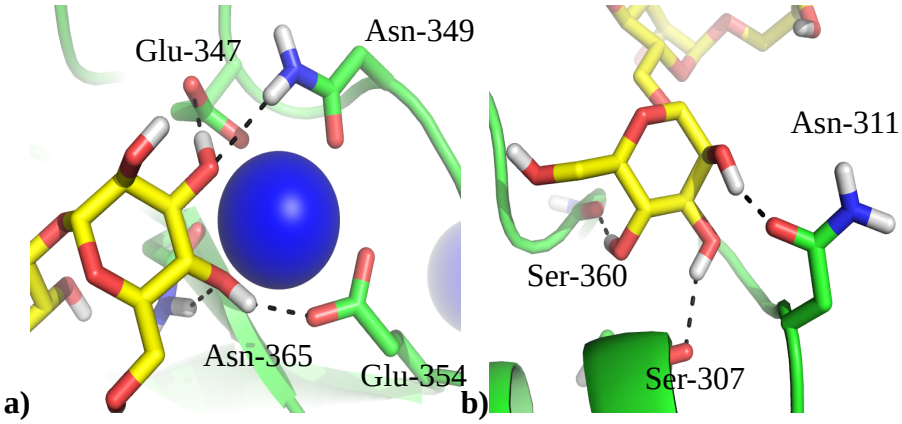

Figura 28: Detalhes das ligações de hidrogênio formadas pelos resíduos **a)** Man-6 e **b)** Man-8 com DC-SIGN. A proteína está mostrada em *cartoon* verde com cadeias laterais em *sticks*, o fragmento de glicosilação em *sticks* amarelos e o Ca^{2+} principal como uma esfera azul.

Podemos decompor o número de ligações de hidrogênio que cada resíduo de manose estabelece com o DC-SIGN ainda mais ao discriminarmos os resíduos da lectina que estão envolvidos na interação. Essa análise revela uma característica interessante do mecanismo molecular de reconhecimento da glicosilação *high-mannose* pelo DC-SIGN: mesmo os resíduos de carboidrato que interagem consistentemente com o DC-SIGN, Man-6 e Man-8, na verdade fazem múltiplas ligações de hidrogênio transientes em vez de poucas ligações que se mantém por todo o tempo de simulação (Tabelas 3 e 4).

Tabela 3: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_5 e cada resíduo do DC-SIGN. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados.

Resíduo	Número de ligações de hidrogênio
man-1	
LYS-373	0.1 ± 0.1
man-5	
GLU-358	0.9 ± 0.1
SER-360	0.3 ± 0.2
man-6	
GLU-347	1.0 ± 0.0
ASN-365	0.7 ± 0.0
ASP-367	0.4 ± 0.3
GLU-354	0.2 ± 0.1
LYS-368	0.1 ± 0.1
ASN-349	0.1 ± 0.0

Resíduo	Número de ligações de hidrogênio
man-2	
GLN-274	0.2 ± 0.2
ALA-372	0.1 ± 0.1
man-8	
SER-360	0.4 ± 0.2
ASN-344	0.4 ± 0.3
ASN-311	0.3 ± 0.3
ASN-362	0.3 ± 0.1
GLY-361	0.2 ± 0.2
ARG-312	0.1 ± 0.1
ARG-345	0.1 ± 0.1
SER-307	0.1 ± 0.1

Tabela 4: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_9 e cada resíduo do DC-SIGN. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados.

Resíduo	Número de ligações de hidrogênio
man-2	
PHE-313	0.2 ± 0.2
man-4	
ASN-344	0.2 ± 0.2
ARG-345	0.1 ± 0.1
man-6	
GLU-347	1.0 ± 0.0
ASN-365	0.5 ± 0.1
ASP-367	0.4 ± 0.3
ASN-349	0.4 ± 0.1
GLU-354	0.3 ± 0.2
LYS-368	0.1 ± 0.1
man-9	
SER-360	0.2 ± 0.2
ARG-312	0.2 ± 0.2
SER-308	0.2 ± 0.1
ASN-311	0.1 ± 0.1

Resíduo	Número de ligações de hidrogênio
man-3	
ASN-311	0.1 ± 0.1
man-5	
GLU-358	0.4 ± 0.3
SER-360	0.1 ± 0.0
man-8	
ASN-344	0.8 ± 0.4
ASN-362	0.7 ± 0.1
GLU-358	0.6 ± 0.6
ARG-345	0.2 ± 0.2

Por fim, analisamos a decomposição das ligações de hidrogênio entre DC-SIGN e os resíduos da lectina em vez do açúcar a fim de identificar os resíduos do DC-SIGN envolvidos nas principais ligações de hidrogênio com a glicosilação *high-mannose*. Para isso foi calculada a ocupância de ligação de hidrogênio para cada resíduo do receptor, ou seja, a porcentagem do tempo de simulação analisado no qual aquele resíduo faz pelo menos uma ligação de hidrogênio com seu ligante.

Foram detectados quatro resíduos que estabelecem ligações de hidrogênio com Man_5 de forma consistente ao longo das simulações: Glu-347, Glu-358, Ser-360 e Asn-365 (Tabela 5). Os quatro resíduos identificados estão localizados em uma mesma região, sendo que dois deles, Glu-347 e Asn-365, fazem parte do sítio de complexação principal de Ca^{2+} (Fig. 29). Também foram detectados quatro resíduos do DC-SIGN que estão consistentemente envolvidos em ligações de hidrogênio com o fragmento da glicosilação maior, Man_9: Glu-347, Asn-344, Glu-358 e Asn-362. Entre eles, apenas Glu-347 faz parte do sítio de complexação principal de Ca^{2+} (Tabela 6).

Além dos resíduos citados, há muitos outros envolvidos no reconhecimento da glicosilação. No entanto, a interação deles com o açúcar é apenas

transiente durante as interações, ou seja, suas interações se formam e se rompem continuamente. Apesar de sua natureza transiente, o grande número dessas interações pode contribuir apreciavelmente para a afinidade e está de acordo com a visão de que carboidratos interagem com seus receptores através de múltiplas interações de baixa afinidade intrínseca[165].

Tabela 5: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com o fragmento Man_5. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho.

Resíduo	Ocupância [%]
GLU-347	**98.91 ± 0.07**
GLU-358	**90 ± 9**
SER-360	**68 ± 11**
ASN-365	**67 ± 1**
ASP-367	36 ± 28
ASN-344	36 ± 22
ASN-311	27 ± 18
ASN-362	24 ± 12
GLN-274	21 ± 18
GLU-354	20 ± 10
GLY-361	17 ± 17
LYS-373	12 ± 10
ALA-372	12 ± 9
ARG-312	10 ± 7
LYS-368	8 ± 8
ASN-349	8 ± 1
ARG-345	7 ± 5
SER-307	7 ± 6
GLY-363	2 ± 2
PHE-313	1 ± 1
PHE-374	0.6 ± 0.6
SER-308	0.4 ± 0.4

Tabela 6: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com o fragmento Man_9. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho.

Resíduo	Ocupância [%]
GLU-347	**96 ± 1**
ASN-344	**69 ± 15**
GLU-358	**68 ± 27**
ASN-362	**67 ± 9**
ASN-365	47 ± 11
ASN-349	40 ± 14
ASP-367	39 ± 30
GLU-354	35 ± 19
SER-360	34 ± 19
ARG-345	28 ± 15
ASN-311	23 ± 2
PHE-313	18 ± 18
ARG-312	18 ± 18
SER-308	16 ± 14
LYS-373	7 ± 7
LYS-368	6 ± 6
GLN-274	5 ± 5
ALA-372	4 ± 4
GLY-361	4 ± 4
GLY-363	1.0 ± 0.8

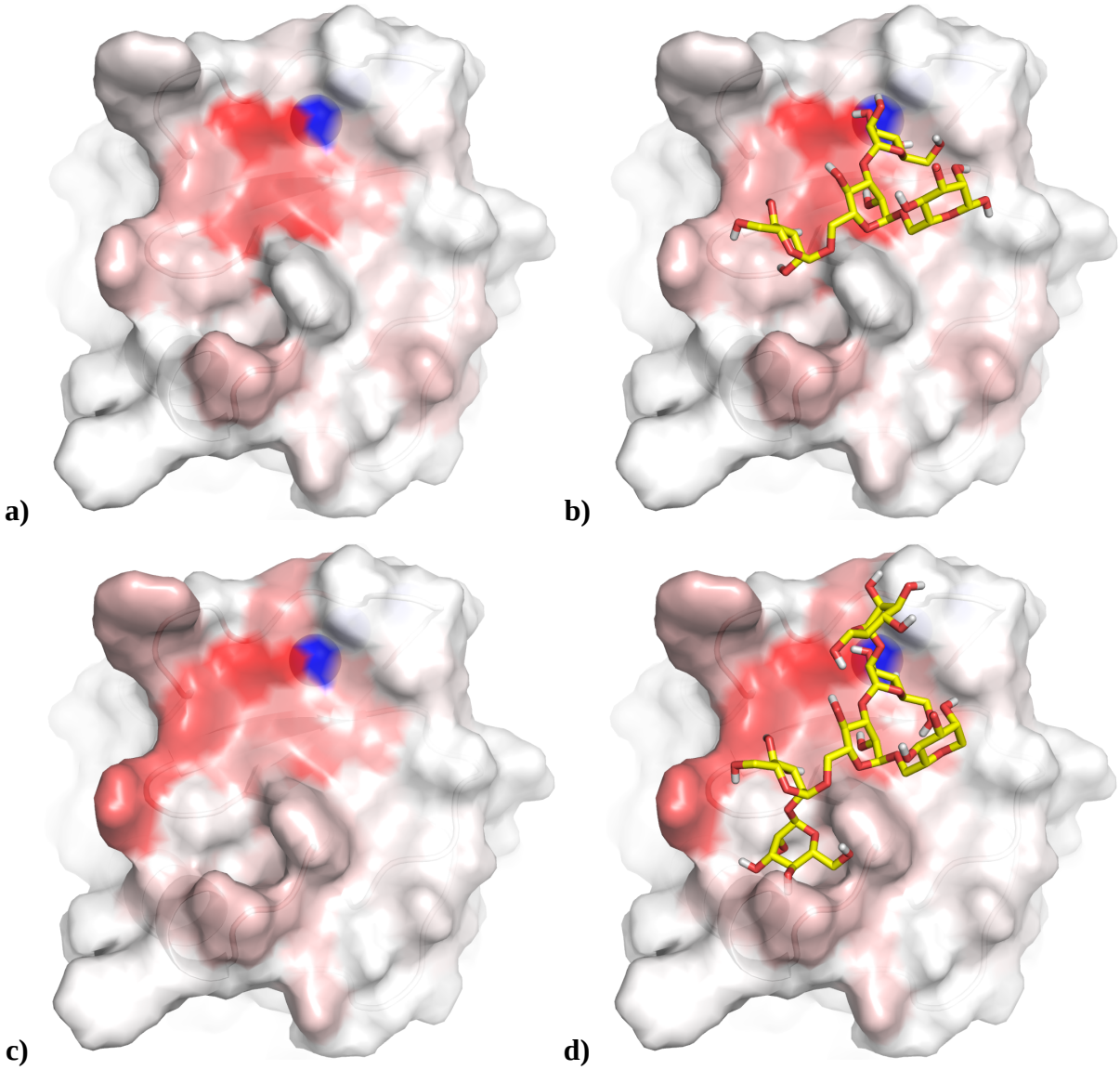

Figura 29: Ocupância das ligações de hidrogênio de cada resíduo do DC-SIGN com os fragmentos da glicosilação *high-mannose* **(a,b)** Man_5 e **(c,d)** Man_9 mostrada em escala de cor gradual de branco, que corresponde a ocupância 0 %, até vermelho, que corresponde a ocupância 100%. A esfera azul é o Ca^{2+} principal. A coluna da direita mostra o mesmo da esquerda com a adição do fragmento de glicosilação **(b)** Man_5 ou **(d)** Man_9, mostrado como *sticks* amarelos.

Como foi visto, foi possível descrever o mecanismo molecular de interação entre DC-SIGN e a glicosilação *high-mannose* através de simulações de dinâmica molecular. As diferenças no envolvimento dos resíduos que reconhecem Man_5 e Man_9 revela a capacidade que o receptor tem de distinguir glicosilações muito similares e, portanto, seu potencial de induzir respostas celulares diferentes a cada um deles.

4.2.2. MR

Assim como feito para o DC-SIGN, o valor de área enterrada de cada um dos resíduos do MR foi calculada a fim de encontrar eventuais interações hidrofóbicas ou de empilhamento de anéis que possam ser importantes para o reconhecimento da glicosilação *high-mannose*. Cinco resíduos destacam-se nessa análise (Fig. 30a): *i*) Leu-694 com o maior valor de área enterrada, 44 ± 3 $Å^2$. É interessante notar que a posição desse resíduo é equivalente a onde está a Phe-313 no DC-SIGN, que também foi identificada como sendo um importante ponto de interação. *ii*) Ile-749 com 10.6 ± 0.7 $Å^2$ de área enterrada. Apesar de esse valor ser pequeno em termos absolutos, ele tem o maior valor de porcentagem de área previamente exposta ao solvente que foi enterrada pelo ligante, 76 ± 2 %. Esse resíduo se localiza nas imediações do sítio de complexação de Ca^{2+}; *iii*) His-753 com 30 ± 5 $Å^2$ de área enterrada, está localizado logo ao lado do resíduo Ile-749; *iv*) His-692 com 28 ± 5 $Å^2$ de área enterrada. É o mais distante do Ca^{2+} principal; e *v*) Tyr-729 com 18 ± 5 $Å^2$ de área enterrada. O envolvimento dos resíduos Tyr-729, Ile-749 e His-753 no reconhecimento de manose já foi descrito anteriormente através de ressonância magnética nuclear[88]. O fato de os mesmos resíduos terem sido detectados como importantes pontos de interação hidrofóbica durante nossas simulações de dinâmica molecular serve como validação do modelo estrutural que construímos combinando as estruturas cristalográficas do DC-SIGN complexando Man_4 com a do MR, que foi obtida na ausência de ligante e na qual foi observado o artefato de troca de domínios (Fig. 30b).

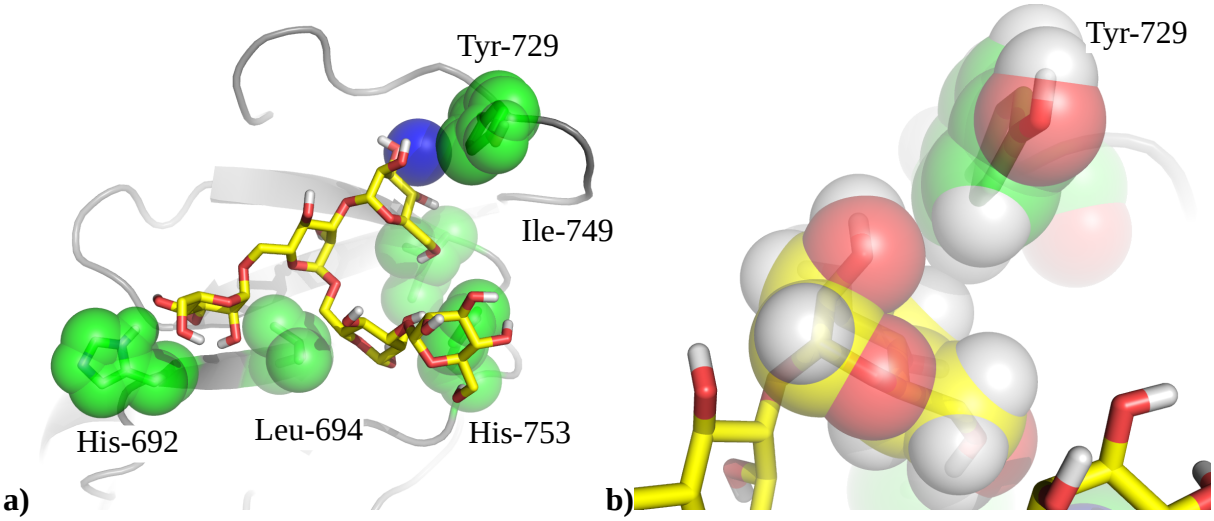

Figura 30: a) Interação entre o fragmento Man_5, representado como *sticks* amarelos, com os resíduos His-692, Leu-694, Tyr-729, Ile-749 e His-753 do MR, destacados como *sticks* e esferas verdes. A proteína está mostrada em *cartoon* cinza e o Ca^{2+} principal como uma esfera azul. **b)** Detalhe do contato entre os anéis da manose principal do Man_5, em amarelo, e Tyr-729 do MR, em verde.

Durante as simulações foram observadas, em média, 4.5 ± 0.6 ligações de hidrogênio entre MR e Man_5 e 8 ± 1 entre o receptor e Man_9. Esses números são semelhantes aos obtidos para o DC-SIGN. A decomposição do total de ligações de hidrogênio em função do envolvimento de cada um dos monossacarídeos dos carboidratos (Tabela 7) mostra que, assim como para o DC-SIGN, menos da metade dos monossacarídeos da glicosilação fazem ligações de hidrogênio de forma consistente com o receptor ao longo das simulações, a saber: Resíduos de manose numerados como Man-5, -6, -8 e -9 (Fig. 31, 32).

Tabela 7: Número médio de ligações de hidrogênio estabelecidas entre MR e cada um dos resíduos de manose dos fragmentos Man_5 e Man_9 ao longo das simulações.

Man-x	Man_5	Man_9
1	0.04 ± 0.03	0.0003 ± 0.0001
2	0.4 ± 0.2	0.6 ± 0.6
3		0.1 ± 0.1
4		0.4 ± 0.3
5	0.7 ± 0.3	0.7 ± 0.3
6	2.4 ± 0.2	2.8 ± 0.1
7		0.09 ± 0.02
8	0.9 ± 0.4	2 ± 1
9		1.4 ± 0.8

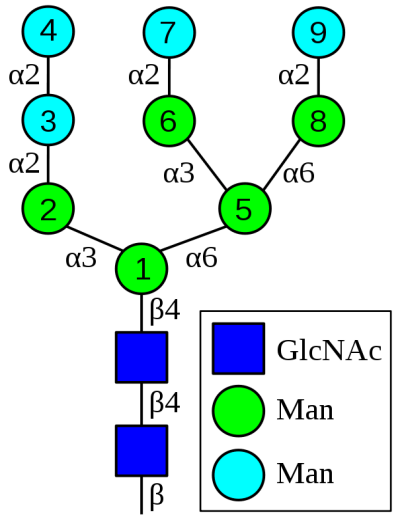

Figura 31: Estrutura esquemática da glicosilação $GlcNAc_2Man_9$ e a numeração de cada resíduo de manose que adotamos. $GlcNAc_2Man_5$ não tem as manoses 3, 4, 7 e 9, destacadas em ciano.

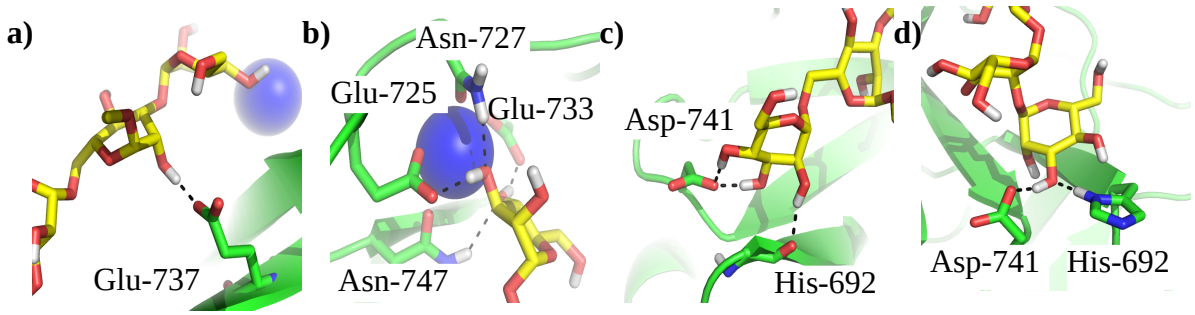

Figura 32: Detalhes das ligações de hidrogênio formadas pelos resíduos **a)** Man-5, **b)** Man-6, **c)** Man-8 e **d)** Man-9 das glicosilações *high-mannose*.

Novamente de forma análoga ao DC-SIGN, a decomposição do número de ligações de hidrogênio que cada monossacarídeo da glicosilação faz em função da contribuição de cada resíduo do MR mostra que o reconhecimento ocorre através de múltiplas

ligações de hidrogênio transientes em vez de poucas que se mantém consistentemente ao longo de toda a simulação (Tabelas 8 e 9).

Tabela 8: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_5 e cada resíduo do MR. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados.

Resíduo	Número de ligações de hidrogênio
man-2	
LEU-754	0.2 ± 0.2
LYS-652	0.1 ± 0.1
HIS-753	0.1 ± 0.0
man-6	
GLU-733	0.9 ± 0.0
ASN-747	0.7 ± 0.1
GLU-725	0.6 ± 0.1
ASN-727	0.1 ± 0.0
man-5	
GLU-737	0.7 ± 0.2
man-8	
HIS-692	0.4 ± 0.1
ASP-741	0.3 ± 0.3
TYR-723	0.1 ± 0.0
LYS-739	0.1 ± 0.0

Tabela 9: Número médio de ligações de hidrogênio estabelecidas entre cada resíduo de manose do fragmento Man_9 e cada resíduo do MR. Todos os resíduos para o qual o número médio é maior ou igual a 0.1 estão mostrados.

Resíduo	Número de ligações de hidrogênio
man-2	
LEU-754	0.3 ± 0.3
ASN-755	0.3 ± 0.3
man-4	
LYS-652	0.1 ± 0.1
LEU-754	0.1 ± 0.1
LYS-649	0.1 ± 0.1
TYR-723	0.1 ± 0.1
man-6	
GLU-725	0.9 ± 0.0
GLU-733	0.9 ± 0.1
ASN-747	0.6 ± 0.2
ASN-727	0.4 ± 0.2
man-7	
TYR-729	0.1 ± 0.0
man-3	
LYS-649	0.1 ± 0.1
man-5	
GLU-737	0.6 ± 0.3
LYS-739	0.1 ± 0.1
man-8	
GLU-737	0.7 ± 0.7
ASN-747	0.3 ± 0.3
HIS-692	0.3 ± 0.3
LYS-739	0.1 ± 0.1
ASP-741	0.1 ± 0.0
man-9	
ASP-741	1.0 ± 0.6
HIS-692	0.4 ± 0.3

A fim de identificar os principais resíduos do MR que interagem com os fragmentos da glicosilação high-mannose por ligação de hidrogênio, a ocupância das ligações de hidrogênio para cada um de seus resíduos foi calculada. Foram identificados quatro resíduos que estão consistentemente envolvidos em ligações de hidrogênio com Man_5 e Man_9 durante as simulações, a saber: Glu-733, Asn-747, Glu-737 e Glu-725 (Tabelas 10 e 11). Os quatro resíduos principais identificados estão em uma mesma região da proteína e Glu-733, Asn-747 e Glu-725 fazem parte do sítio de complexação de Ca^{2+} (Fig. 33). Ao contrário do encontrado para o DC-SIGN, o mecanismo de reconhecimento das glicosilações de glicosilações *high-mannose* do MR não é sensível às diferenças entre os fragmentos Man_5 e Man_9.

Tabela 10: Ocupância das ligações de hidrogênio de cada resíduo do MR com o fragmento Man_5. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho.

Resíduo	Ocupância [%]
GLU-733	**92 ± 3**
ASN-747	**67 ± 8**
GLU-737	**65 ± 24**
GLU-725	**63 ± 12**
HIS-692	39 ± 6
ASP-741	20 ± 17
LEU-754	19 ± 16
ASN-727	13 ± 4
HIS-753	13 ± 4
LYS-739	10 ± 2
TYR-723	7 ± 4
LYS-652	7 ± 5
ASN-731	3 ± 2
THR-743	2 ± 2
LYS-649	2 ± 1
SER-745	2 ± 2
ASN-750	0.5 ± 0.5
ASN-755	0.3 ± 0.2
LYS-693	0.2 ± 0.2
GLN-730	0.2 ± 0.1
LEU-694	0.1 ± 0.1
HIS-650	0.01 ± 0.01

Tabela 11: Ocupância das ligações de hidrogênio de cada resíduo do MR com o fragmento Man_9. Todos os valores acima de zero estão mostrados. As médias acima de 50% estão destacadas em vermelho.

Resíduo	Ocupância [%]
GLU-725	**93 ± 5**
GLU-737	**92 ± 8**
GLU-733	**86 ± 8**
ASN-747	**85 ± 7**
ASP-741	69 ± 25
HIS-692	60 ± 25
LEU-754	39 ± 29
ASN-727	39 ± 15
ASN-755	31 ± 31
LYS-739	27 ± 11
LYS-649	14 ± 7
LYS-652	10 ± 10
TYR-723	8 ± 5
TYR-729	5.3 ± 0.6
GLN-730	3 ± 2
SER-745	2 ± 1
HIS-753	2 ± 2
HIS-650	2 ± 2
ASN-750	1 ± 1
ASN-731	1 ± 1
THR-743	0.7 ± 0.4
LYS-693	0.5 ± 0.5
GLY-740	0.4 ± 0.2
GLU-752	0.2 ± 0.2

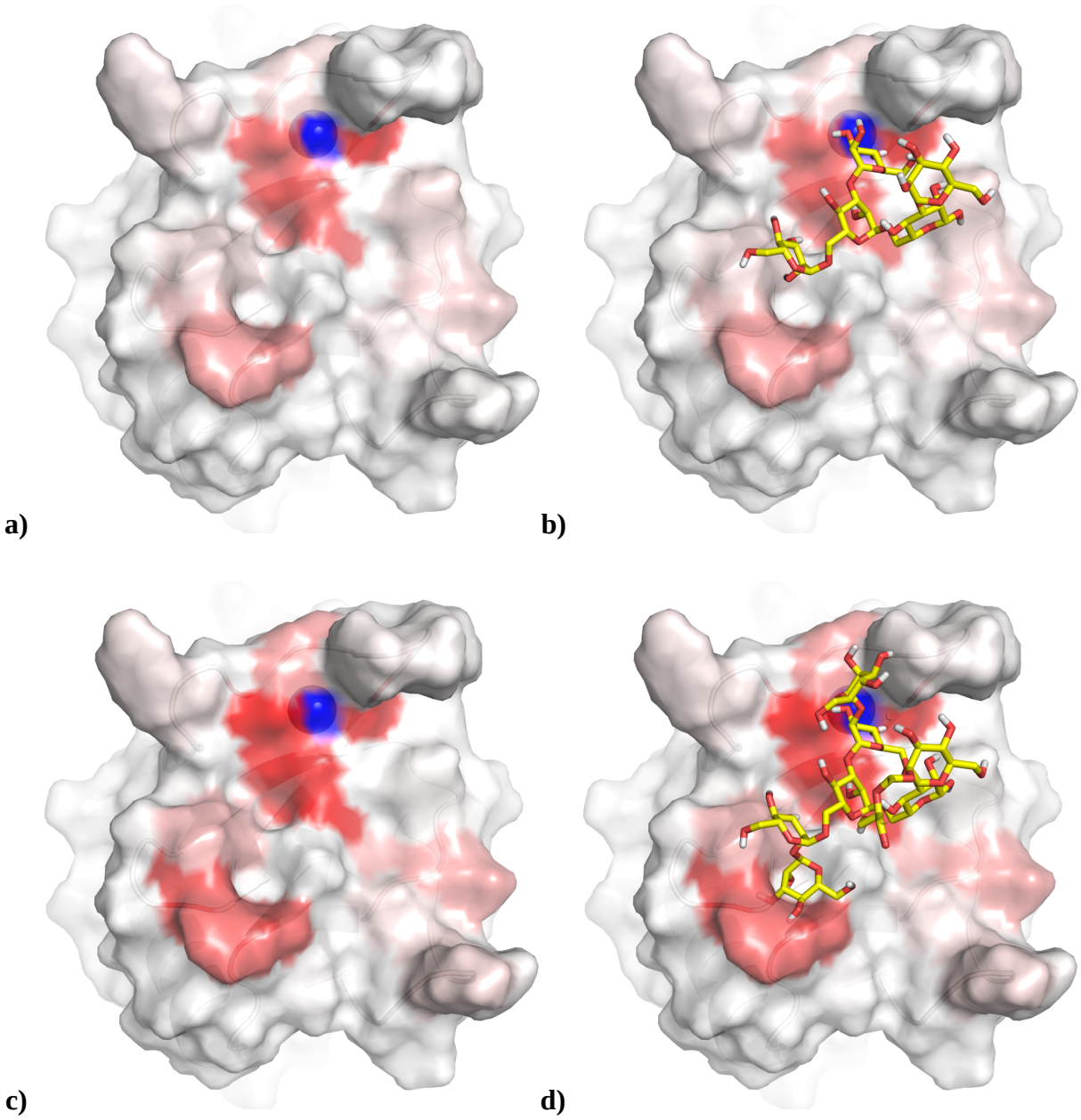

Figura 33: Ocupância das ligações de hidrogênio de cada resíduo do MR com os fragmentos da glicosilação *high-mannose* **(a,b)** Man_5 e **(c,d)** Man_9 mostrada em escala de cor gradual de branco, que corresponde a ocupância 0 %, até vermelho, que corresponde a ocupância 100%. A esfera azul é o Ca^{2+} principal. A coluna da direita mostra o mesmo da esquerda com a adição do fragmento de glicosilação **(b)** Man_5 ou **(d)** Man_9, mostrado como *sticks* amarelos.

4.3. Rotâmeros da glicosilação *high-mannose*

Como último passo na caracterização do mecanismo de reconhecimento de carboidratos das CTLs, construímos modelos contendo as glicosilações *high-mannose* inteiras, $GlcNAc_2Man_5$ e $GlcNAc_2Man_9$. Isso foi feito para o DC-SIGN a partir da estrutura cristalográfica na qual ele complexa o fragmento Man_4 (PDB ID 1SL4). Então procedemos ao

cálculo de energia livre de interação entre DC-SIGN e essas glicosilações *high-mannose* completas através do método *Umbrella Sampling*. O fragmento Le^x, que representa as glicosilações *complex-type*, foi incluído nessa análise.

Após testes exploratórios iniciais, escolhemos usar como a coordenada de reação espacial ξ a distância entre o centro de massa do CRD do DC-SIGN e 1-OH, a hidroxila na posição 1 do primeiro monossacarídeo de cada glicosilação. O centro de massa do carboidrato não foi usado porque pode flutuar devido ao pequeno tamanho e alta flexibilidade do açúcar. Adicionalmente, 1-OH é o átomo que estaria ligado ao vírus da Dengue no sistema biológico. Portanto nossa coordenada de reação reflete a distância lectina–vírus esperada. A amostragem dessa coordenada de reação foi feita com simulações de dinâmica molecular dirigidas em que os dois grupos supracitados foram separados por um potencial harmônico que se move, o que também pode ser visualizado como uma mola cuja distância de equilíbrio aumenta com o tempo de simulação.

Frames foram selecionados a partir dessas trajetórias em intervalos regulares de 1 Å na coordenada ξ desde a menor distância amostrada até 5.5 nm, distância na qual consideramos não haver mais interação receptor–ligante. Essa seleção resulta em aproximadamente 40 janelas de simulação. Cada janela foi submetida a termalização de 1 ns e produção de 50 ns no qual um potencial harmônico em repouso foi usado para restringir a amostragem de ξ. Nos casos em que não houve boa sobreposição dos histogramas de amostragem repetimos as simulações necessárias para melhorar a amostragem usando estruturas iniciais diferentes.

O PMF (do inglês, "*Potential of Mean Force*") mostra o perfil de energia livre ao longo da coordenada de reação. Uma vez que estamos lidando com ligantes grandes e flexíveis, primeiro analisamos a convergência do PMF calculado. Para a interação entre o CRD do DC-SIGN e a glicosilação $GlcNAc_2Man_5$ o sistema relaxa após 20 ns de simulação e mantém um comportamento consistente até 50 ns. O ΔG calculado, que é a diferença entre o menor valor do PMF e seu valor a uma distância na qual define-se que não há interação, muda de aproximadamente -12 kcal/mol durante os primeiros 20 ns para aproximadamente -9 kcal/mol depois desse ponto. Adicionalmente, o perfil do PMF sugere que o processo de ligação ocorre sem estados intermediários (Fig. 34a).

Apesar de $GlcNAc_2Man_9$ ser um açúcar maior, ele relaxa depois de apenas 10 ns para um comportamento consistente. No entanto, a extensão da diferença observada é maior: o ΔG vai de -12 kcal/mol antes do relaxamento para -6 kcal/mol depois de 10 ns. Curiosamente, neste caso o perfil do PMF indica que há vários estados intermediários ao

longo da coordenada de reação. Isso indica que os resíduos de manose adicionais interagem extensivamente com o receptor (Fig. 34b).

Já a glicosilação *complex-type* Le^x mostra um comportamento diferente. Seu PMF indica que há oscilação entre alguns valores de ΔG que vão de -3 a -6 kcal/mol (Fig. 34c). Isso indica que estamos amostrando pelo menos três estados conformacionais diferentes. No entanto, convenientemente, seu PMF não muda significativamente quando mais de um desses estados é considerado simultaneamente (Fig. 34c).

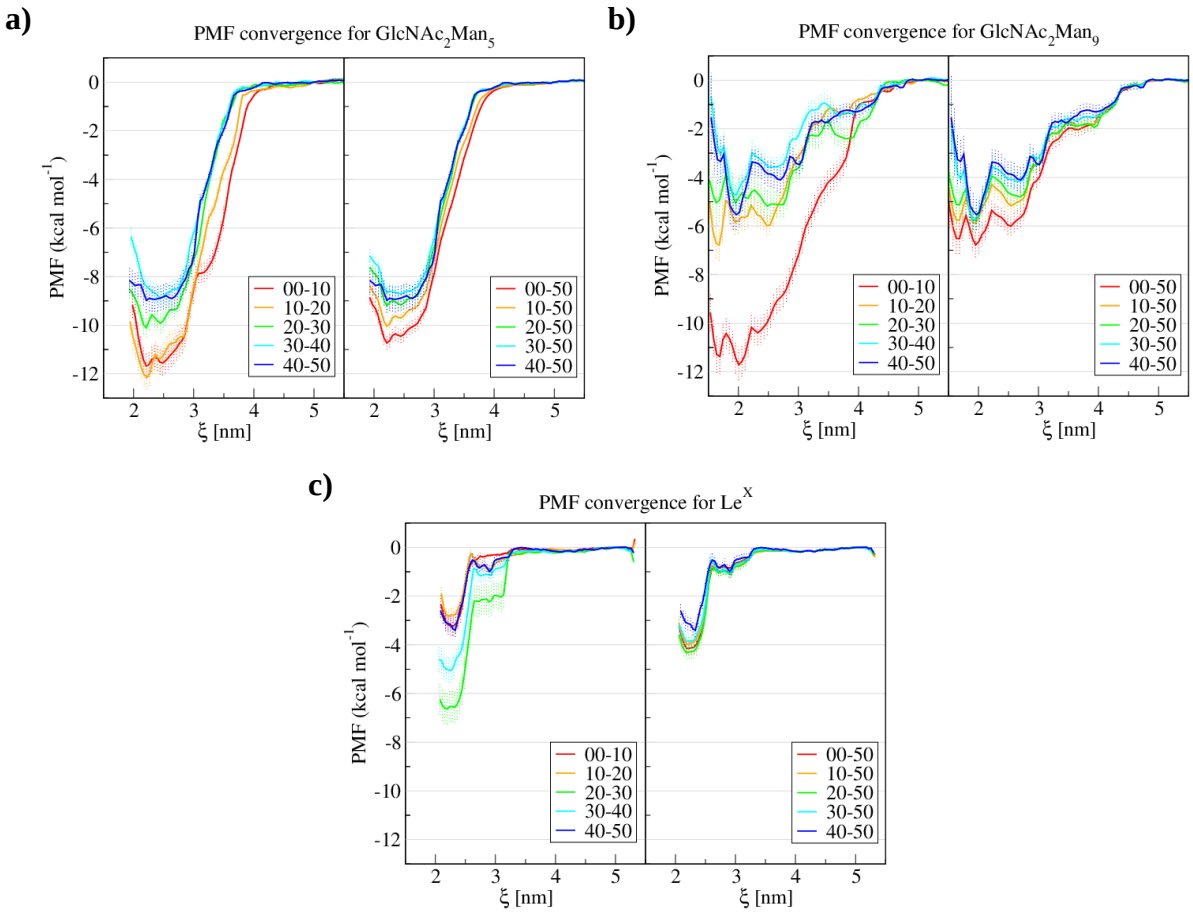

Figura 34: PMFs calculados considerando vários intervalos das simulações de *Umbrella Sampling* para a interação entre o CRD do DC-SIGN e **a)** $GlcNAc_2Man_5$; **b)** $GlcNAc_2Man_9$; **c)** Le^x. Para cada complexo são mostrados dois gráficos: à esquerda são mostrados PMFs calculados considerando intervalos de 10 ns que não se sobrepõem. À direita são mostrados PMFs calculados a partir de um determinado tempo que varia de 0 até 40 ns em passos de 10 ns até o fim da simulação, 50 ns.

Assim, consideramos o intervalo de 20 a 50 ns das simulações de *Umbrella Sampling* para calcular o ΔG de interação (Fig. 35a). Os histogramas de amostragem para os três sistemas mostraram boa sobreposição (Fig. 35b-d).

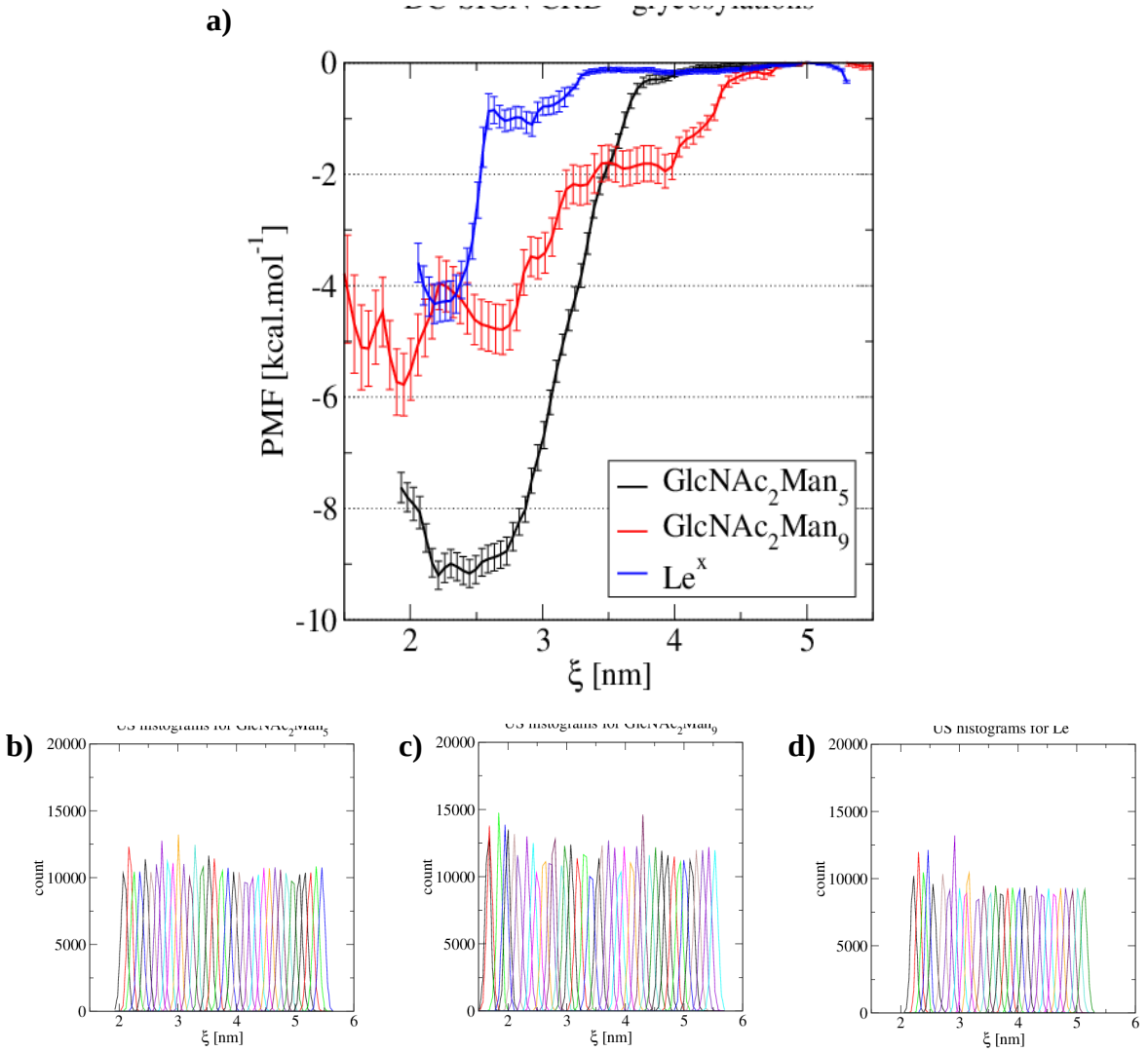

Figura 35: a) PMF calculado pelo método *Umbrella Sampling* para descrever a interação entre o CRD do DC-SIGN e as glicosilações $GlcNAc_2Man_5$, $GlcNAc_2Man_9$ e Le^x. Histogramas de amostragem para as glicosilações: **b)** $GlcNAc_2Man_5$, **c)** $GlcNAc_2Man_9$ e **d)** Le^x.

Os valores de ΔG de interação obtidos foram -9.2±0.3, -5.8±0.6 e -4.3±0.4 kcal/mol para $GlcNAc_2Man_5$, $GlcNAc_2Man_9$ and Le^X, respectivamente. No entanto, esses valores não estão de acordo com o comportamento esperado com base na afinidade relativa das glicosilações pela lectina em um experimento de *glycan array*[75]. De acordo com os dados experimentais de referência, $GlcNAc_2Man_9$ se liga aproximadamente 1.5 vezes mais ao DC-SIGN do que Le^X, o que levaria a um $\Delta\Delta G$ esperado de aproximadamente 0.30 kcal/mol em favor do açúcar *high-mannose*. Nosso $\Delta\Delta G$ observado de quase 1.5 kcal/mol, apesar de qualitativamente correto, está quantitativamente errado. Essa discrepância pode ser explicada por diferenças na apresentação desses açúcares no arranjo experimental. Foi observado que a forma através da qual cada açúcar é ligado à superfície do *chip* no *glycan array* pode ter um grande impacto sobre as afinidades medidas. Especificamente, a afinidade do DC-SIGN pelo

açúcar Le^a, que é similar a Le^x, aumenta em quase três vezes quando um *linker* mais longo é usado[75]. As propriedades de ligação a superfície e apresentação do Le^x e $GlcNAc_2Man_9$ são diferentes devido a seus tamanhos diferentes; Por isso, não sabemos a extensão do efeito do *linker* para comparação quantitativa com o experimento.

Por outro lado, as propriedades de apresentação e ligação a superfícies de $GlcNAc_2Man_5$ e $GlcNAc_2Man_9$ são similares devido a seu tamanho e estrutura similares. De acordo com os dados de *glycan array*, $GlcNAc_2Man_9$ se liga aproximadamente dez vezes mais ao CRD do DC-SIGN do que $GlcNAc_2Man_5$[75]. Disso podemos estimar um $\Delta\Delta G$ de aproximadamente 1.4 kcal/mol em favor do açúcar maior. No entanto, observamos $\Delta\Delta G$ de aproximadamente 3 kcal/mol em favor do açúcar menor. Esse resultado foi reproduzido em vários *setups* de simulação e análise diferentes. Por isso, buscamos possíveis falhas no modelo construído a partir da estrutura cristalográfica.

Um aspecto importante que não foi considerado previamente é a variabilidade conformacional das glicosilações *high-mannose*. De forma interessante, o fragmento de glicosilação que está presente na estrutura cristalográfica não corresponde a nenhum dos seus rotâmeros mais estáveis. O ângulo diedro ω da Man-1 corresponde ao rotâmero gg mas o mesmo ângulo da Man-5 está numa conformação eclipsada com ω de aproximadamente 20º.

Para considerar a variabilidade conformacional na nossa descrição das glicosilações *high-mannose* pelo DC-SIGN, geramos estruturas 3D otimizadas das glicosilações $GlcNAc_2Man_5$ e $GlcNAc_2Man_9$ para todos os rotâmeros usando o servidor GLYCAM Web. Então, com o objetivo de considerar a possibilidade de outras manoses além da Man-5 interagirem no sítio principal, alinhamos cada manose de cada rotâmero àquela que está interagindo com o Ca^{2+} principal na estrutura cristalográfica de referência (PDB ID 1SL4). Isso gera um total de 20 possíveis complexos para $GlcNAc_2Man_5$ e 36 para $GlcNAc_2Man_9$. Todos esses 56 sistemas foram submetidos à minimização de energia e à termalização de 1 ns e produção de 50 ns com 5 réplicas. A maioria das 56 possibilidades não é fisicamente factível devido a sobreposição de átomos da glicosilação e da proteína. Especificamente, para $GlcNAc_2Man_5$ 11 das 20 possibilidades, cerca de 55%, não tem sobreposições e para $GlcNAc_2Man_9$ isso ocorre para apenas 14 das 36 possibilidades, cerca de 39%.

Anteriormente foi mostrada a ocupância de ligações de hidrogênio entre o CRD do DC-SIGN e os fragmentos Man_5 e Man_9. Agora estendemos essa abordagem para os complexos entre o receptor e todas as possíveis conformações das glicosilações completas. O perfil de ocupância de ligação de hidrogênio dos complexos construídos a partir da estrutura cristalográfica mostra que há alta ocupância alta principalmente perto do Ca^{2+} principal, como esperado com base no mecanismo de interação conhecido. Adicionalmente, alguns outros resíduos têm ocupância menor, indicando que suas interações são transientes (Fig. 36a,e).

Surpreendentemente, alguns rotâmeros mostram comportamento marcadamente diferente uns dos outros. Há casos em que a ocupância de ligações de hidrogênio vai de interações localizadas com alguns poucos resíduos em diferentes partes da proteína até interações que se espalham pelas imediações do sítio de ligação de Ca^{2+} (Fig. 36). Apenas um resíduo tem alta ocupância de forma consistente em todas as conformações: Glu-347, que é um dos resíduos complexando o Ca^{2+} principal. Esses resultados reforçam nossa descrição anterior de que o DC-SIGN é um receptor capaz de distinguir diferenças estruturais nos carboidratos que complexa. Demonstramos que o mecanismo de reconhecimento de açúcares é sensível a diferenças mínimas, como é o caso de diferentes rotâmeros de uma mesma glicosilação.

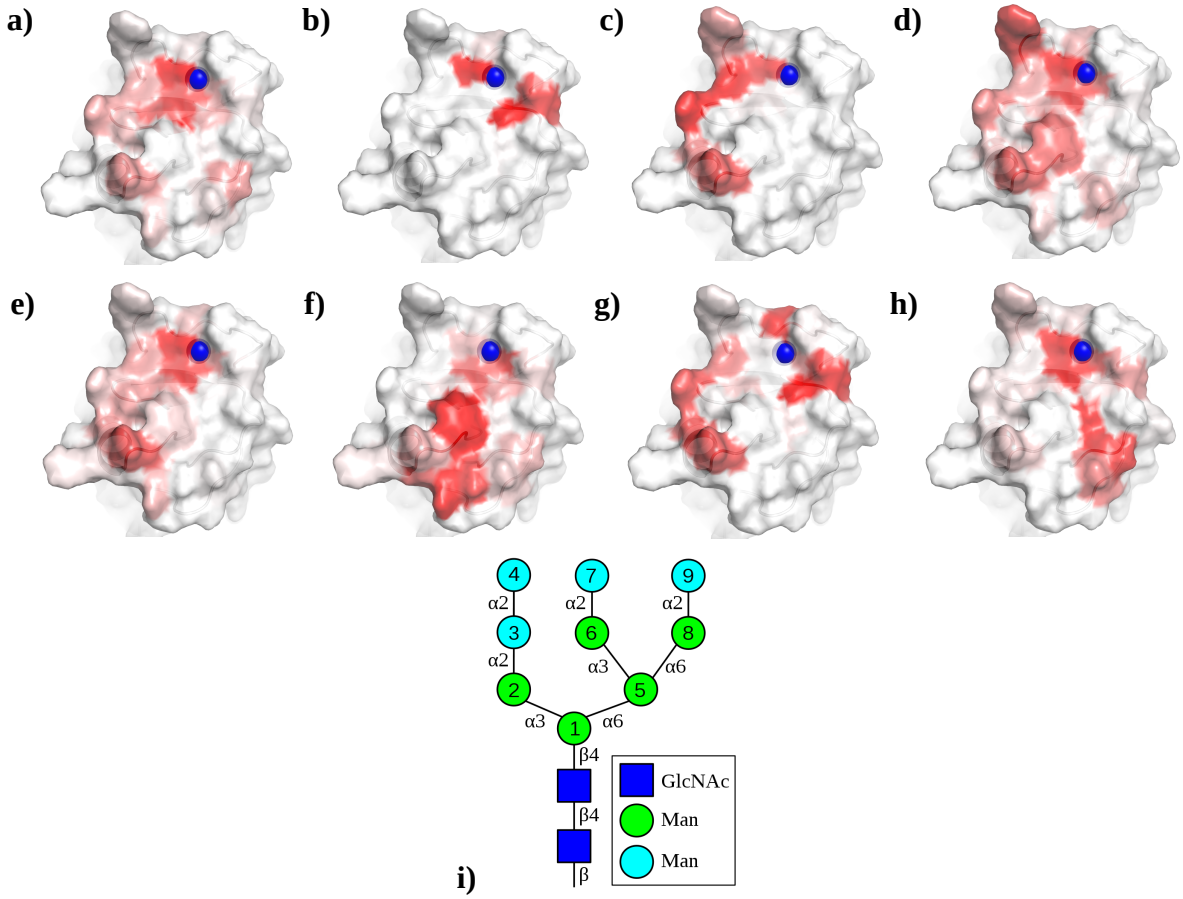

Figura 36: Ocupância das ligações de hidrogênio entre DC-SIGN e as glicosilações *high-mannose* projetada sobre sua estrutura 3D. A escala de cores vai de branco a vermelho para indicar 0 e 100% de ocupância, respectivamente. A esfera azul é o Ca^{2+} principal. **Linha superior:** $GlcNAc_2Man_5$, sendo **a)** conformação construída a partir da estrutura cristalográfica de referência; **b)** rotâmero gt-gg interagindo com o Ca^{2+} principal pela Man-1; **c)** rotâmero gt-gt interagindo com o Ca^{2+} principal pela Man-3; **d)** rotâmero gg-gt interagindo com o Ca^{2+} principal pela Man-7. **Linha central:** $GlcNAc_2Man_9$, sendo **e)** conformação construída a apartir da estrutura cristalográfica de referência; **f)** rotâmero gg-gg interagindo com o Ca^{2+} principal pela Man-4; **g)** rotâmero gg-gt interagindo com o Ca^{2+} principal pela Man-8; **h)** rotâmero gt-gt interagindo com o Ca^{2+} principal pela Man-4. **i)** Numeração dos resíduos de manose da glicosilação.

Uma vez que o mecanismo de reconhecimento dos rotâmeros muda qualitativamente, isso é, pelos resíduos específicos que estão envolvidos, e quantitativamente, isso é, pela ocupância com os resíduos interagentes, podemos esperar que cada uma dessas conformações tenha uma afinidade diferente pelo receptor. Para calcular tais afinidades, idealmente calcularíamos o ΔG entre o CRD do DC-SIGN CRD e cada conformação das glicosilações usando *Umbrella Sampling* da mesma forma que foi feito para os sistemas derivados da estrutura cristalográfica. No entanto, o custo computacional dessa abordagem é proibitivo. Portanto, focamos nossos recursos no cálculo das afinidades relativas dos rotâmeros da glicosilação $GlcNac_2Man_5$ segundo três diferentes métodos de menor custo computacional. Especificamente, usamos MM/PBSA e igualdade de Jarzynski para calcular ΔG e simulações dirigidas para calcular a força de ruptura, que está correlacionada com ΔG (Fig. 37).

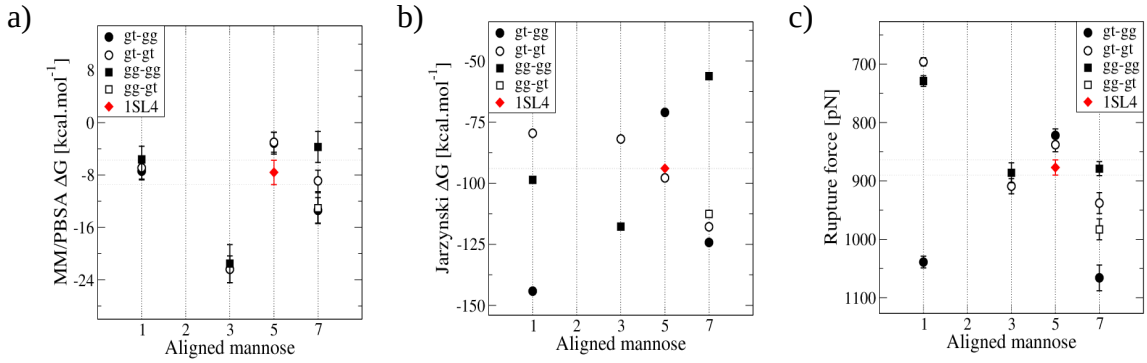

Figura 37: Métricas usadas para estimar o ΔG relativo de interação para cada rotâmero da glicosilação $GlcNAc_2Man_5$: **a)** MM/PBSA; **b)** Igualmente de Jarzynski, as barras de erro são menores do que os símbolos do gráfico; **c)** Força de ruptura, o eixo y está invertido para facilitar a comparação com os outros gráficos. O valor obtido para o caso construído com base na estrutura cristalográfica (PDB ID 1SL4) está destacado em vermelho.

Os únicos rotâmeros que tem melhor afinidade pelo CRD do DC-SIGN do que a estrutura construída com base na estrutura cristalográfica de acordo com todas as métricas, isso é, força de ruptura maior e menor ΔG para MM/PBSA e igualdade de Jarzynski, são gt-gg e gg-gt quando interagindo com o Ca^{2+} principal pela sua Man-7. Adicionalmente, o rotâmero gt-gt alinhado por Man-1 e gt-gg alinhado por Man-5 tem consistentemente menos afinidade pelo receptor do que o sistema de referência. O comportamento dos demais rotâmeros é inconsistente entre as três métricas.

Uma vez que o comportamento dos sistemas foi inconsistente, faz-se necessário discutir as diferenças entre esses métodos. O método MM/PBSA usa solvatação implícita e eletrostática contínua para calcular a energia livre de solvatação. O principal limitação dessa abordagem é que a contribuição de moléculas de água do solvente para o mecanismo de reconhecimento e energia de ligação são ignorados. Portanto, em situações nas quais água tem um papel importante como ponte entre o receptor e o ligante, MM/PBSA vai

inferir ligação mais fraca do que realmente é. Vale destacar que o ΔG de ligação é calculado de acordo com a estrutura cristalográfica, -8±2 kcal/mol, concorda bem com o valor obtido por *Umbrella Sampling*, -9.2±0.3 kcal/mol. O maior erro estatístico no valor do método MM/PBSA pode ser reduzido pelo uso de um número maior de *frames* de simulação nos cálculos.

Uma vez que os cálculos de força de ruptura e aqueles usando a igualdade de Jarzynski usam o mesmo tipo de simulações dirigidas, todos os possíveis problemas associados à coordenada de reação espacial afetam ambos igualmente. A diferença chave entre esses métodos é que o valor da força de ruptura depende apenas da maior barreira energética no processo de rompimento do complexo enquanto o ΔG calculado via igualdade de Jarzynski contém informação sobre todas as barreiras. Em casos em que apenas uma interação receptor–ligante domina o processo de complexação, podemos esperar que ambas os métodos se comportem de maneira similar. No entanto, se há várias interações fracas sem dominância clara, a força de ruptura vai inferir ligação mais fraca do que a igualdade de Jarzynski. O fator limitante para aplicar a igualdade de Jarzynski é o número de simulações independentes necessárias para o cálculo de ΔG absoluto.

4.4. O comportamento da igualdade de Jarzynski

A igualdade de Jarzynski diz que a diferença de energia livre de equilíbrio pode ser calculada a partir de uma distribuição de trabalho obtida pelo processo fora do equilíbrio desde que um número suficientemente grande de realizações independentes sejam consideradas. No limite de um processo de *pulling* infinitamente lento, precisaríamos de apenas uma trajetória porque o negativo da diferença de energia livre de Gibbs de interação é, por definição, o trabalho necessário para induzir o processo de desligamento. No entanto, o custo computacional de tal abordagem pode ser muito grande. Adicionalmente, a ineficiência de amostragem pode exigir que essa simulação extremamente longa seja repetida algumas vezes. A igualdade de Jarzynski permite que consigamos o mesmo resultado fazendo um número maior de simulações de *pulling* mais rápidas.

Para cada realização da simulação de *pulling* fora do equilíbrio, podemos calcular o trabalho que a mola virtual fez sobre o sistema para separar completamente o carboidrato do receptor e, assim, construir a distribuição de trabalho (Fig. 38a). Essa distribuição pode ser usada para estimar a energia livre de interação através da igualdade de Jarzynski. Observamos que o PMF calculado desta forma se parece muito com o da trajetória de menor trabalho incluída no cálculo (Fig. 38b, linha vermelha). Isso acontece como consequência da média exponencial; trajetórias com menores valores de trabalho recebem peso estatístico muito maior do que aquelas nas quais a mola fez mais trabalho (Fig. 38b).

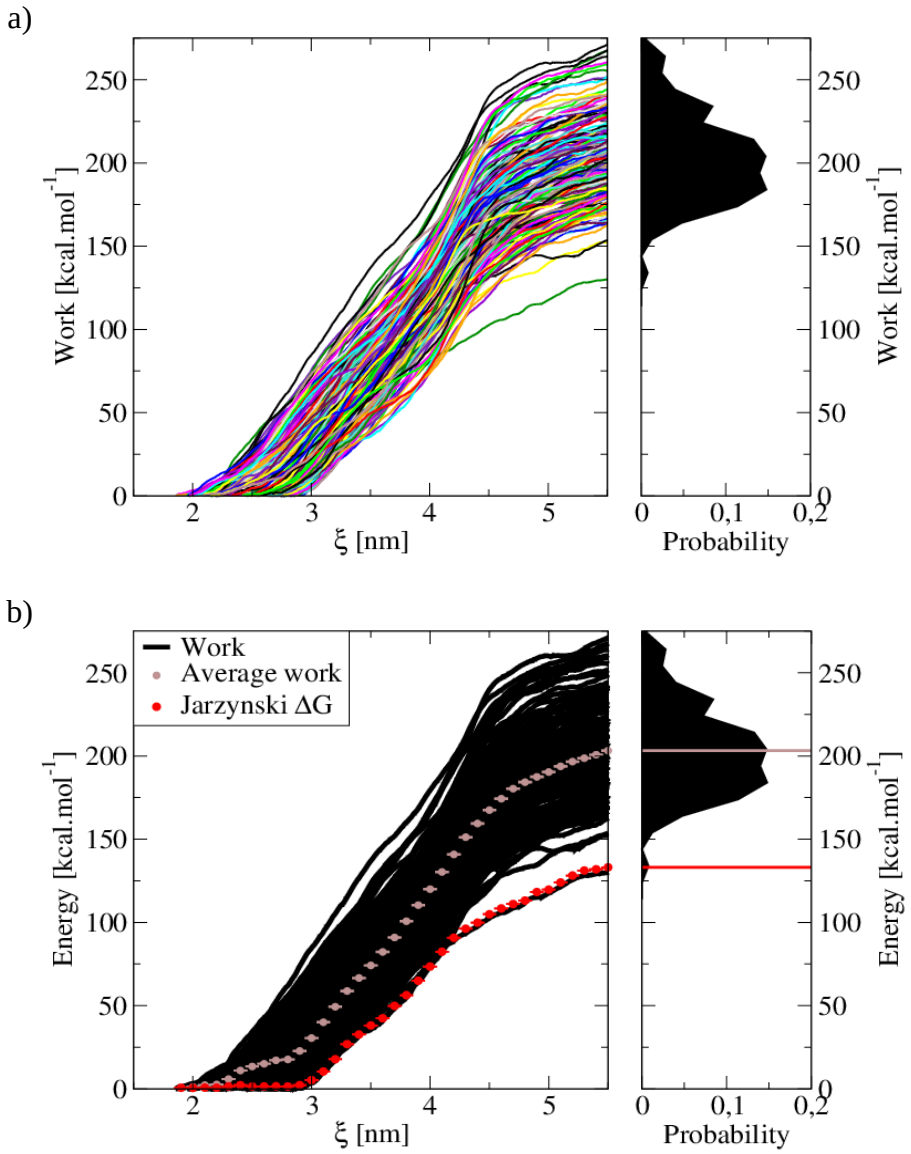

Figura 38: Comportamento das 165 trajetórias de simulação dirigidas feitas com v = 200 Å/ns e k = 1000 pN/Å. **a)** Perfis e distribuição de trabalho; **b)** Trabalho médio e estimativa de diferença de energia livre de acordo com a igualdade de Jarzynski.

A seguir, analisamos o comportamento da estimativa de energia livre de acordo com a igualdade de Jarzynski variando a velocidade de *pulling*, constante de força da mola e *ensemble* termodinâmico (Tabela 12). Enquanto o *ensemble* e constante da mola não influenciaram a distribuição de trabalho, a velocidade de *pulling* afetou profundamente os resultados, como esperado. Especificamente, a distribuição de trabalho foi deslocada para menores valores conforme a velocidade de *pulling* diminuiu. A estimativa de ΔG seguiu a mesma tendência: para as velocidades de *pulling* de 200, 100, 10 e 1 Å/ns a estimativa de ΔG é -119.009±0.002, -86.770±0.002, -26.923±0.003 e -17.752±0.001 kcal/mol, respectivamente (Fig. 39).

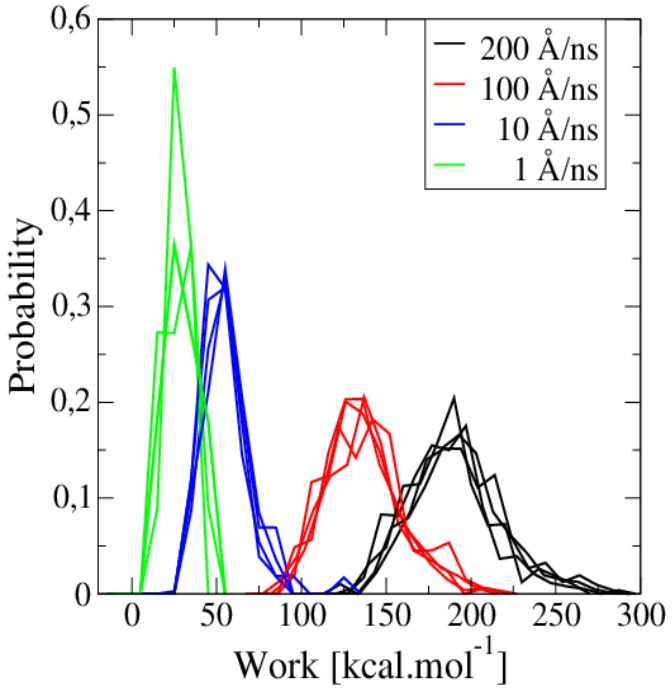

Figura 39: Distribuições de trabalho calculadas de simulações de pulling com *v* = 200 Å/ns (preto), 100 Å/ns (vermlho), 10 Å/ns (azul) e 1 Å/ns (verde). Simulações feitas com *k*=166 pN/Å, *k*=1000 pN/Å, *ensemble* NPT e *ensemble* NVT estão mostradas na mesmas cores.

Tabela 12: Parâmetros para todas as simulações dirigidas fora do equilíbrio feitas com o sistema que descreve a interação entre o CRD do DC-SIGN e a glicosilação $GlcNAc_2Man_5$ construída conforme a estrutura cristalográfica de referência.

Velocidade de pulling, *v* [Å/ns]	Constante de força da mola, *k* [pN/Å]	*Ensemble* termodinâmico	Repetições	Tempo total de simulação [ns]
200	1000	NPT	165 x 150ps	25
200	1000	NVT	165 x 150ps	25
200	166	NPT	165 x 150ps	25
200	166	NPT	165 x 150ps	25
100	1000	NPT	165 x 300ps	50
100	1000	NVT	165 x 300ps	50
100	166	NPT	165 x 300ps	50
100	166	NPT	165 x 300ps	50
10	1000	NPT	55 x 3000ps	165
10	1000	NVT	55 x 3000ps	165
10	166	NPT	55 x 3000ps	165
10	166	NPT	55 x 3000ps	165
1	1000	NPT	11 x 30000ps	330
1	1000	NVT	11 x 30000ps	330
1	166	NPT	11 x 30000ps	330
1	166	NPT	11 x 30000ps	330

Em vez de executar simulações com uma velocidade de *pulling* ainda menor, como 0.1 Å/ns, aumentamos a amostragem para a velocidade v =1 Å/ns com constante de força k = 1000 pn/Å devido ao baixo número de trajetórias inicialmente consideradas. O aumento do número de trajetórias independentes de 11 para 123 muda a estimativa de energia livre de -17.752±0.001 para -9.111±0.003 kcal/mol (Fig. 40), em perfeita concordância com os cálculos de *Umbrella Sampling* de ΔG = -9.2±0.3 kcal/mol.

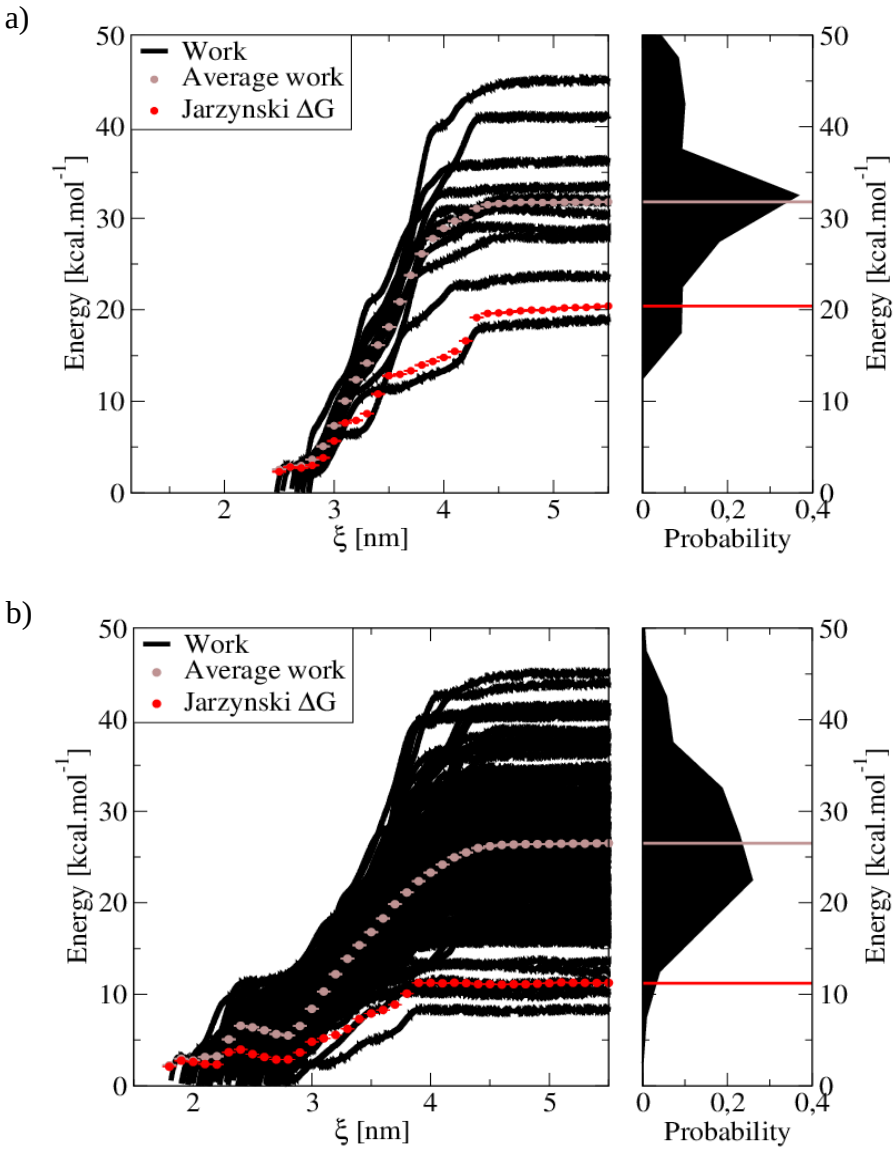

Figura 40: Comportamento das simulações de *pulling* feitas com v = 1 Å/ns e k = 1000 pN/Å. **a)** amostragem de 11 trajetórias ; **b)** amostragem de 123 trajetórias.

A comparação dos PMFs obtidos por *Umbrella Sampling* e iguadade de Jarzynski para o caso com v=1 Å/ns e k=1000 pN/Å mostra que o caminho que os métodos amostraram do estado complexado até o dissociado é diferente. Enquanto no PMF calculado com *Umbrella Sampling* há apenas um mínimo que aumenta monotonamente até zero, aquele calculado com igualdade de Jarzynski mostra dois mínimos com incrementos até zero em três

passos, indicando que pelo menos três interações-chave são rompidas ao longo da coordenada de reação (Fig. 41). A forma geral do PMF não muda quando a amostragem é aumentada de 11 para 123 trajetórias independentes, indicando que o caminho amostrado continua o mesmo.

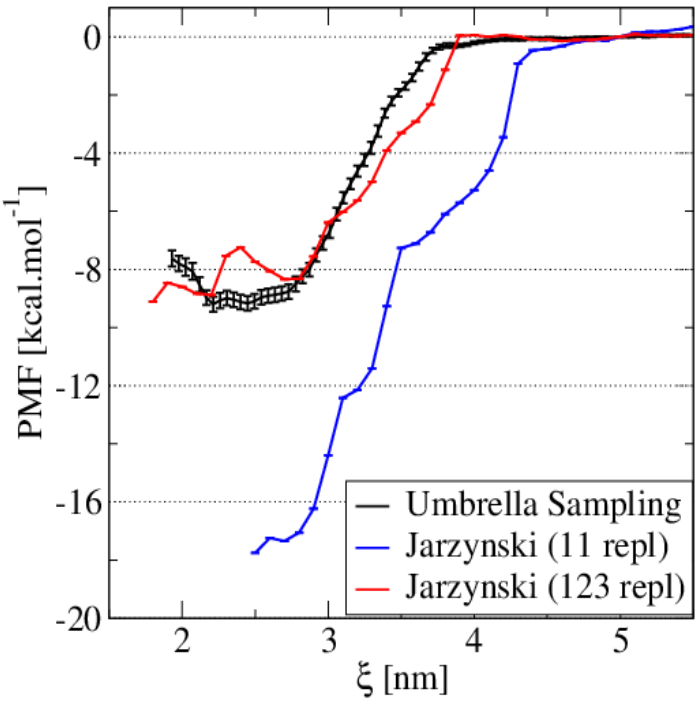

Figura 41: PMFs descrevendo a interação entre o CRD do DC-SIGN e a glicosilação $GlcNAc_2Man_5$ calculados com os métodos Umbrela Sampling (preto), igualdade de Jarzynski com 11 trajetórias independentes com v = 1 Å/ns e k =1000 pN/Å (azul) e igualdade de Jarzynski com 123 trajetórias independentes com v = 1 Å/ns e k =1000 pN/Å (vermelho).

Apesar de ter sido possível obter o mesmo valor de diferença de energia livre de interação através de *Umbrella Sampling* e igualdade de Jarzynski, o custo computacional total deles foi muito diferente. Enquanto cada ΔG calculado por *Umbrella Sampling* exigiu 2.0 µs de tempo total de simulação, considerando 40 janelas de 50 ns cada, a estimativa feita com a igualdade de Jarzynski exigiu quase o dobro usando exatamente a mesma caixa de simulação, 3.7µ ns considerando 123 repetições de 30 ns cada.

Enquanto o número de simulações necessárias para a aplicação do método *Umbrella Sampling* é facilmente estimado *a priori*, a duração das simulações de amostragem dependerá de quanto tempo o sistema demora para atingir o equilíbrio ao longo da coordenada de reação. Já a aplicação da igualdade de Jarzynski nos leva à situação oposta: sabemos *a priori* qual é a duração de cada uma das simulações dirigidas fora do equilíbrio mas não sabemos quantas realizações independentes são necessárias.

4.5. Seleção de moléculas promissoras via ensaios de *virtual screening*

Apesar das milhares ou até milhões de poses geradas em um ensaio de *virtual screening* descreverem razoavelmente bem os possíveis modos de interação dos ligantes com o receptor de interesse, as afinidades relativas das poses pelo receptor não é confiavelmente predita pelas funções de *scoring* usualmente empregadas[102]. Portanto, em vez de usar a função de *scoring* do programa de *docking* adotado —AutoDock Vina— utilizamos duas outras métricas para seleção de moléculas promissoras a partir dos resultados dos ensaios de *virtual screening*: número de ligações de hidrogênio receptor–ligante de cada pose e a área da proteína que é removida do acesso ao solvente na presença do ligante, chamada de área enterrada.

A fim de possibilitar o uso dessas métricas, e uma vez que não há uma ferramenta oficial para execução de ensaios de *virtual screening* com AutoDock Vina sem depender de interfaces gráficas, fizemos uma implementação própria. Nosso conjunto de *scripts* e programas possibilita a preparação dos arquivos necessários, execução do *virtual screening* com o *software* AutoDock Vina e análise dos resultados através do cálculo das métricas citadas. Todos os passos podem ser executados de forma paralela dentro de uma máquina com vários núcleos de processamento ou mesmo através do uso simultâneo de várias máquinas heterogêneas que estão conectadas entre si apenas pela internet.

O número de ligações de hidrogênio e os valores de área enterrada são calculados não só para cada pose como um todo, mas também para cada resíduo do receptor. Ou seja, ao fim da análise dos resultados do *virtual screening* com nossos *scripts* e programas temos, para cada pose: *i*) o número total de ligações de hidrogênio receptor–ligante; *ii*) uma lista contendo cada ligação de hidrogênio individual especificando qual resíduo do receptor está envolvido na interação; *iii*) a área enterrada do receptor; e *iv*) uma lista com a área enterrada de cada resíduo do receptor.

Uma aplicação especialmente interessante da decomposição do número de ligações de hidrogênio e da quantidade de área enterrada em termos dos resíduos do receptor para cada pose é a possibilidade de usar conhecimento prévio sobre o mecanismo molecular de reconhecimento de ligantes pelo receptor de interesse. É possível, por exemplo, selecionar poses que têm alto número de ligações de hidrogênio com um conjunto de resíduos específicos que estão envolvidos no mecanismo de reconhecimento de substratos naturais. A quantidade de área enterrada de resíduos específicos pode ser usada da mesma forma. É possível, ainda, combinar as duas métricas usando conjuntos de resíduos diferentes para construir um *ranking* baseado no conhecimento sobre o mecanismo molecular de interação do receptor com outras moléculas.

Esse conjunto de *scripts* e programas foi usado para executar ensaios de *virtual screening* com o objetivo de encontrar potenciais inibidores da interação entre as glicosilações do vírus da Dengue e as lectinas DC-SIGN e MR. O CLEC5A não foi incluído nesta etapa porque não se dispõe ainda de informações sobre seus possíveis sítios de interação com ligantes.

Para esses ensaios de *virtual screening* usamos uma biblioteca virtual de ligantes que contém moléculas já aprovadas como fármacos pelo FDA. Isso caracteriza um ensaio de reposicionamento de drogas, uma tendência atual em *drug discovery*. Trata-se da tentativa de encontrar novas aplicações terapêuticas para moléculas que já são usadas clinicamente. A principal vantagem é que, uma vez que essas moléculas já foram aprovadas para uso em humanos, suas características de toxicidade e biodisponibilidade já foram caracterizadas[166]. Nossa biblioteca foi construída a partir da junção dos catálogos "*FDA-approved drugs*" e "*drugbank-approved*" da base de dados ZINC[157], totalizando 4888 moléculas.

Além disso, a fim de considerar a flexibilidade do DC-SIGN e MR nos ensaios de *virtual screening*, usamos 60 estruturas extraídas das simulações de dinâmica molecular para cada um deles. As suas estruturas cristalográficas também foram incluídas. O volume de busca do *docking* incluiu toda a região próxima do sítio principal de complexação de Ca^{2+} (Fig. 42).

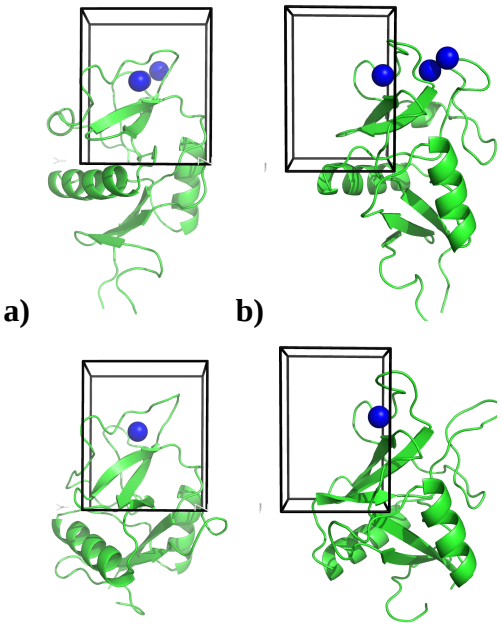

Figura 42: Definição do volume de busca dos *dockings* executados durante o ensaio de *virtual screening*. **Linha superior:** DC-SIGN; **linha inferior:** MR. **Coluna da esquerda:** visão "frontal". **Coluna da direita:** rotação de 90º em relação à coluna da esquerda. A proteína está mostrada como *cartoon* verde, Ca^{2+} como esferas azuis e a caixa que define o volume de busca está destacada com contornos pretos.

4.5.1. DC-SIGN

Iniciamos a seleção de moléculas com alto potencial de interação com DC-SIGN a partir dos ensaios de *virtual screening* através da seleção daquelas com o maior número de ligações de hidrogênio com o receptor. As duas melhores moléculas de acordo com essa métrica fazem 22 ligações de hidrogênio com o receptor. Ambas são aminoglicosídeos usados como antibióticos. Entre elas, a com maior área enterrada é Kanamycin (ZINC08101133) (Fig.43).

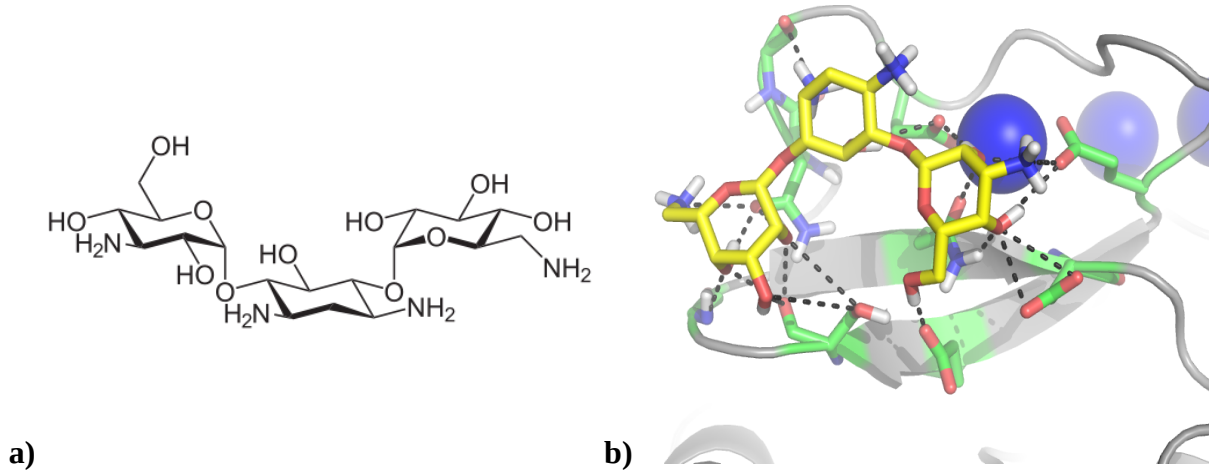

Figura 43: a) Estrutura 2D da Kanamycin. **b)** Interação entre DC-SIGN e Kanamycin. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em sticks verdes. O ligante está mostrado em *sticks* amarelos. Os Ca^{2+} estão mostrados como esferas azuis. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

Uma vez que tais moléculas são um tipo de carboidrato modificado e a abordagem glicomimética já foi explorada na literatura antes[63,64], procuramos outros tipos moléculas que possam ter alto potencial de interação com o DC-SIGN. De forma interessante, as melhores 70 moléculas, que fazem de 22 a 16 ligações de hidrogênio com o DC-SIGN, são carboidratos, análogos deles ou polióis, que é um tipo de molécula semelhante à forma aberta dos carboidratos.

Duas moléculas não-carboidrato finalmente aparecem entre as 80 encontradas formando 15 ligações de hidrogênio com o receptor: Iohexol (ZINC03830943), um agente de contraste para exames de imagem, foi selecionado para a etapa de refinamento por ter maior número de poses diferentes entre as duas moléculas (Fig. 44).

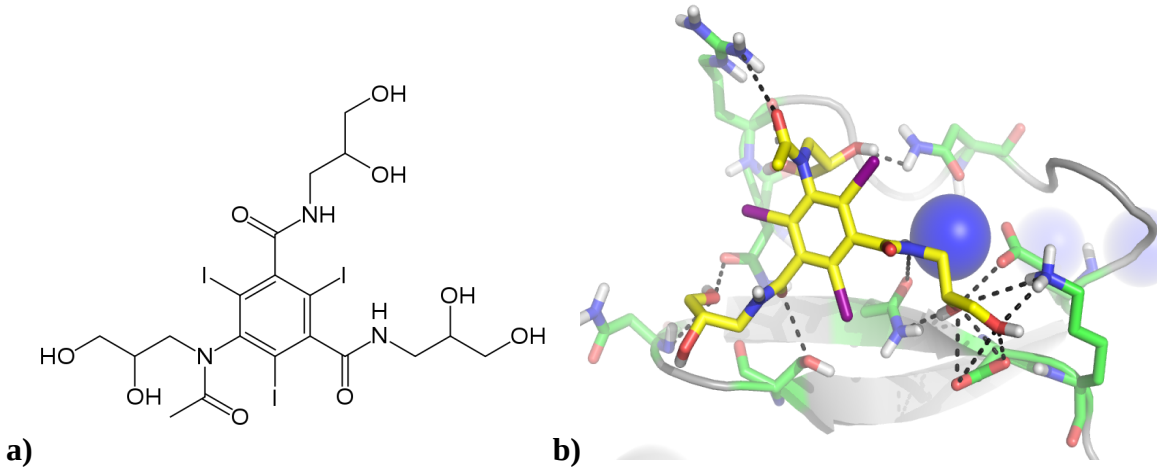

Figura 44: a) Estrutura 2D do Iohexol. **b)** Interação entre DC-SIGN e Iohexol. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

Conforme descrito anteriormente, o sistema implementado permite a seleção de poses com base no número de ligações de hidrogênio com algum conjunto específico de resíduos de interesse. Assim, buscamos as moléculas com o maior número de ligações de hidrogênio especificamente com os resíduos Glu-347, Glu-358, Ser-360 e Asn-365 devido a sua identificação como os resíduos-chave de interação entre DC-SIGN e a glicosilação *high-mannose*. As melhores poses encontradas dessa forma, um total de 20, fazem 9 ligações de hidrogênio com os resíduos selecionados. No entanto, todas correspondem a ligantes que são carboidratos ou derivados de carboidratos, com destaque para os aminoglicosídeos novamente. Prosseguindo na filtragem, foram encontradas 48 poses que fazem 8 ligações de hidrogênio, sendo que a única correspondente a uma molécula diferente de carboidrato foi a Oxytetracycline (ZINC03831260), um antibiótico do grupo das tetraciclinas (Fig. 45). Essa molécula foi selecionada para o passo de refinamento.

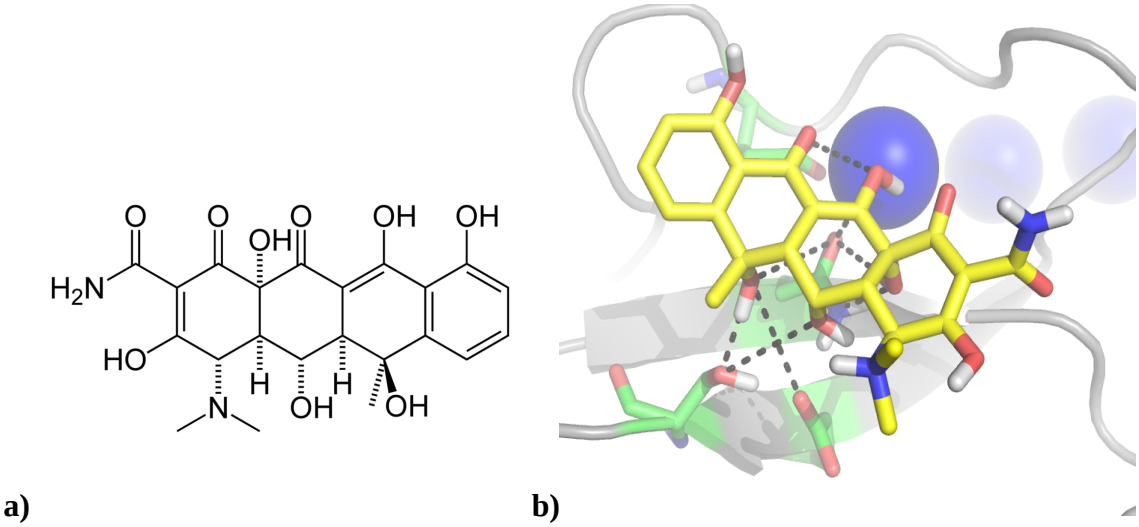

Figura 45: a) Estrutura 2D da Oxytetracycline. **b)** Interação entre DC-SIGN e Oxytetracycline. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

A informação de que o resíduo Phe-313 do DC-SIGN é um importante ponto de interação hidrofóbica com as glicosilações *high-mannose* foi incorporada à filtragem dos resultados do *virtual screening*. Assim, o ligante que enterra a maior área do resíduo Phe-313 ao mesmo tempo em faz ligações de hidrogênio com os quatro resíduos citados anteriormente é Paclitaxel (ZINC96006020), usada contra alguns tipos de câncer. Sua pose enterra 109.4 $Å^2$ do resíduo Phe-313 e faz duas ligações de hidrogênio com os resíduos Glu-347, Glu-358, Ser-360 e Asn-365 (Fig. 46).

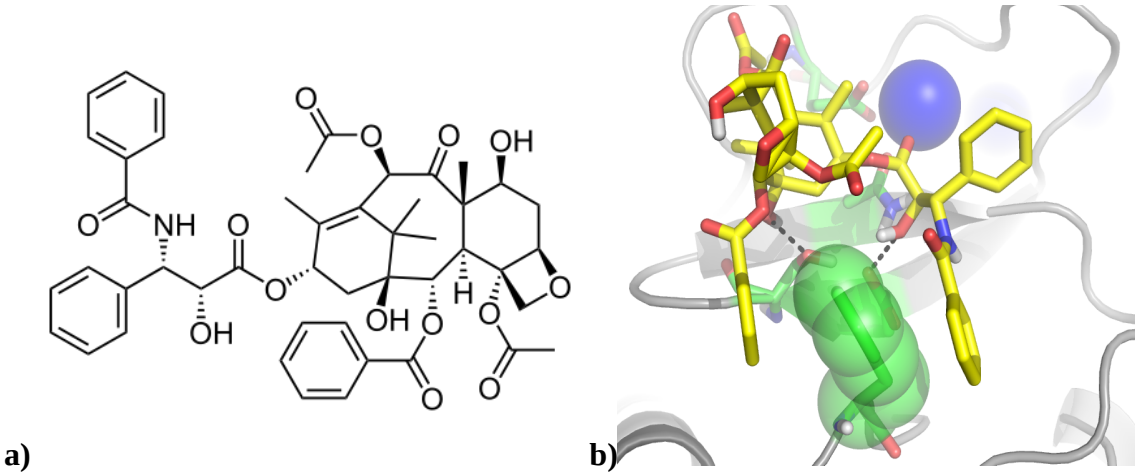

Figura 46: a) Estrutura 2D do Paclitaxel. **b)** Interação entre DC-SIGN e Paclitaxel. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. O resíduo Phe-313 está destacado em esferas verdes transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

Finalmente, como critério de comparação, a melhor molécula de acordo com a função de *score* do AutoDock Vina foi selecionada para a etapa de refinamento: Brown HT (ZINC03830332), um corante marrom usado em alimentos, com afinidade predita pela função de *score* de -9 kcal/mol (Fig. 47).

a) **b)**

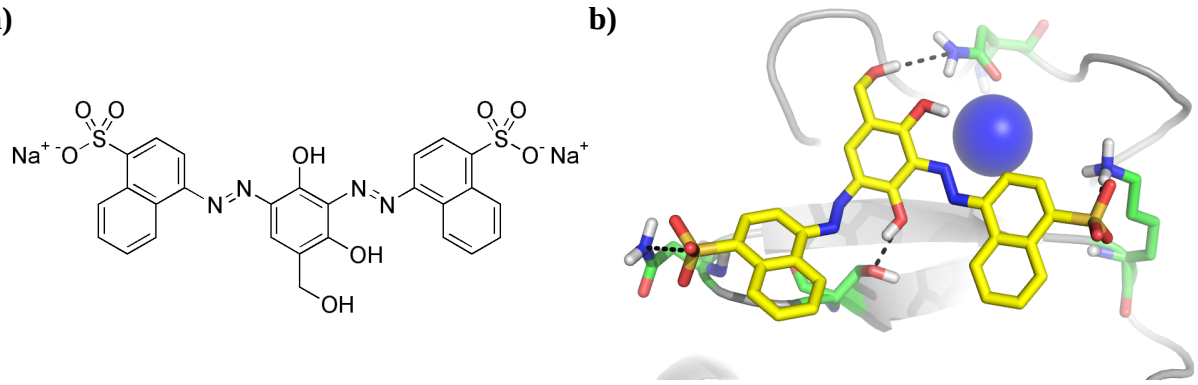

Figura 47: a) Estrutura 2D do Brown HT como um sal de sódio. **b)** Interação entre DC-SIGN e Brown HT. O DC-SIGN está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. Os Ca^{2+} estão mostrados como esferas azuis. O ligante está mostrado em *sticks* amarelos. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

4.5.2. MR

De forma análoga ao feito para o DC-SIGN, iniciamos a seleção de moléculas com alto potencial de interação com o MR pela métrica de número total de ligações de hidrogênio com a proteína. As melhores moléculas encontradas fazem 18 ligações de hidrogênio com o MR. Todas são, novamente, aminoglicosídeos. A melhor delas foi Amikacin (ZINC08101058) (Fig. 48), que também é usada como antibiótico. Todas as 29 moléculas que fazem de 18 a 14 ligações de hidrogênio com o MR são aminoglicosídeos, carboidratos ou polióis.

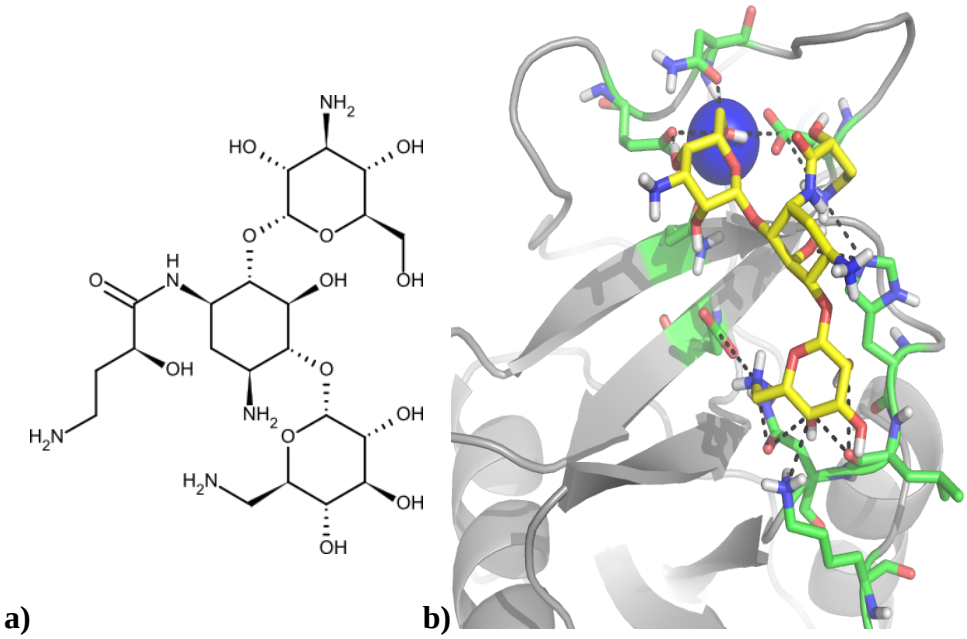

Figura 48: a) Estrutura 2D da Amikacin. **b)** Interação entre MR e Amikacin. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

As melhores moléculas não-carboidrato encontradas fazem 13 ligações de hidrogênio com o receptor. De forma interessante, trata-se exatamente das mesmas encontradas para o DC-SIGN. A molécula Iohexol (ZINC03830943) foi selecionada para a etapa de refinamento novamente (Fig.49).

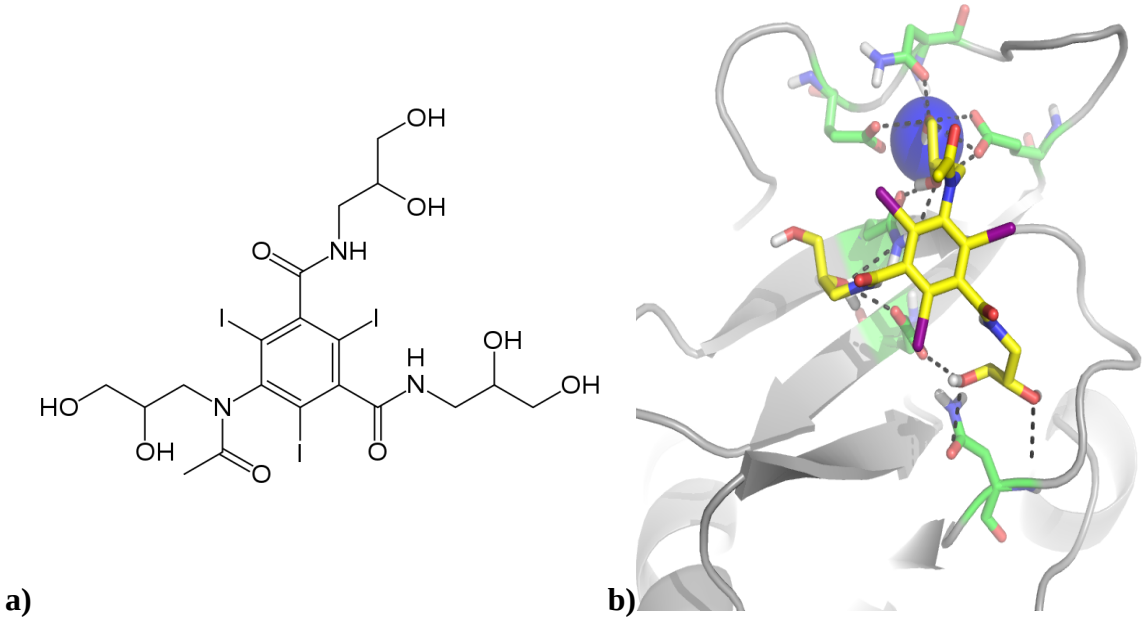

Figura 49: a) Estrutura 2D do Iohexol. **b)** Interação entre MR e Iohexol. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

A seguir limitamos as ligações de hidrogênio aos resíduos Glu-733, Asn-747, Glu-737 e Glu-725, uma vez que, como mostrado anteriormente, eles são importantes para o reconhecimento de carboidratos pelo MR. As melhores moléculas encontradas fazem 9 ligações de hidrogênio com os resíduos selecionados. Todas as 107 poses que fazem 9 ligações de hidrogênio são novamente carboidratos, seus análogos ou polióis. Já entre as 239 poses que apresentam 8 ligações de hidrogênio, há algumas moléculas não-carboidrato; Entre elas o Chloramphenicol (ZINC00113386), que é um antibiótico. Essa molécula é especialmente interessante porque ao mesmo tempo em que tem um alto número de ligações de hidrogênio com os principais resíduos do MR que estão envolvidos no reconhecimento da glicosilação *high-mannose*, também tem uma interação de *stacking* com Tyr-729 (Fig. 50), que também é importante para o reconhecimento de carboidratos do MR.

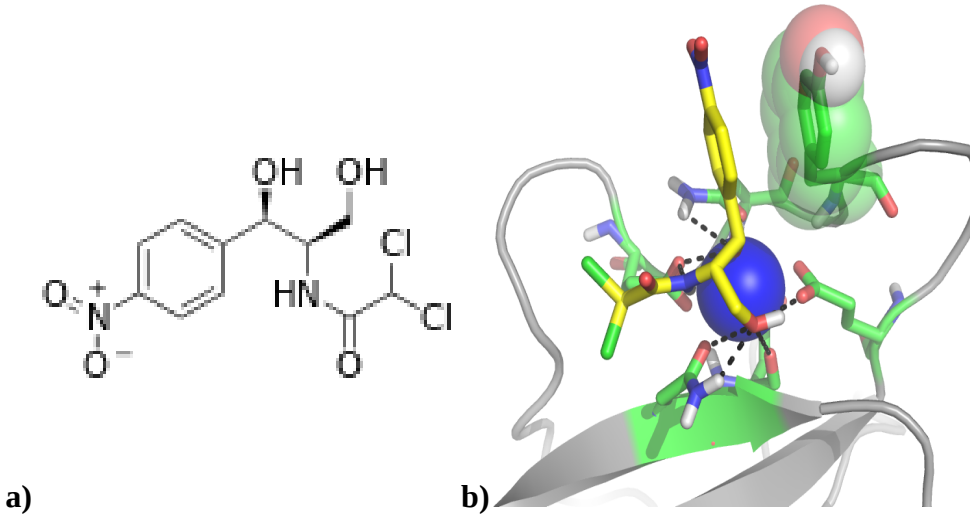

Figura 50: a) Estrutura 2D do Chloramphenicol. **b)** Interação entre MR e Chloramphenicol. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. O resíduo Tyr-729 está destacado em esferas transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

A área do resíduo Leu-694 que é enterrada pela interação com os ligantes foi usada como métrica adicional às ligações de hidrogênio com os quatro resíduos identificados. Esse resíduo foi escolhido porque foi identificado como o principal ponto de interação hidrofóbica para reconhecimento da glicosilação *high-mannose* pelo MR. Assim, foram identificadas duas moléculas que enterram 51.7 $Å^2$ da área da Leu-694 e fazem 5 ligações de hidrogênio com os quatro resíduos selecionados. Uma dessas moléculas foi descartada por ser um surfactante. A outra, Ceftaroline Fosamil (ZINC96006023), um antibiótico da classe das cefalosporinas, foi selecionada para a etapa de refinamento (Fig. 51).

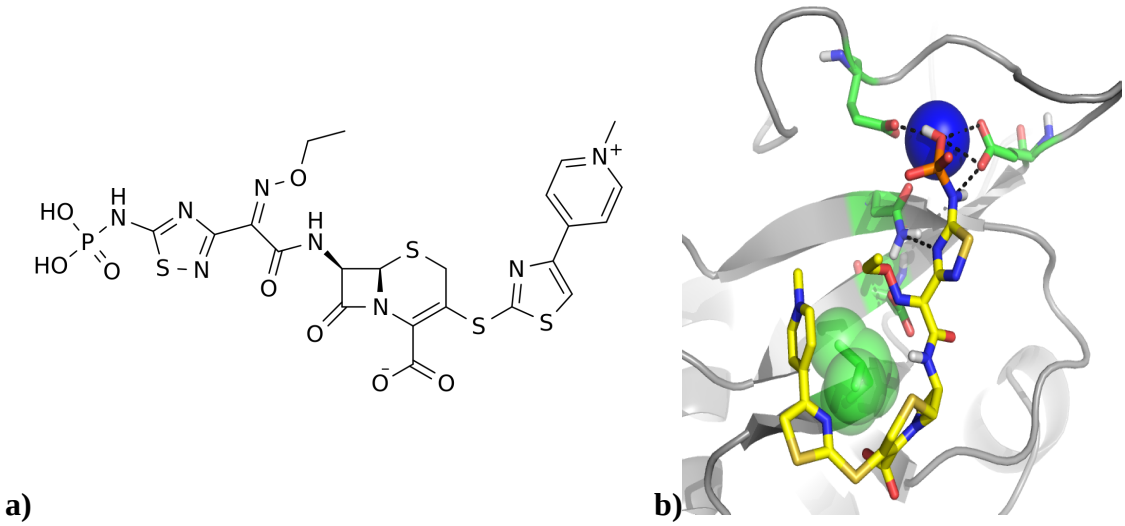

Figura 51: a) Estrutura 2D do Ceftaroline Fosamil. **b)** Interação entre MR e Ceftaroline Fosamil. O MR está mostrado em *cartoon* cinza com os resíduos que participam da interação em *sticks* verdes. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. O resíduo Leu-694 está destacado em esferas transparentes. As ligações de hidrogênio possíveis estão indicadas em pontilhado preto.

Finalmente, para comparação, a melhor molécula de acordo com a função de *score* do AudoDock Vina foi selecionado para a etapa de refinamento. Trata-se da molécula Metocurine (ZINC04198845), um relaxante muscular (Fig. 52). A afinidade predita pela função de *score* é -8.5 kcal/mol que não forma ligações de hidrogênio com o receptor.

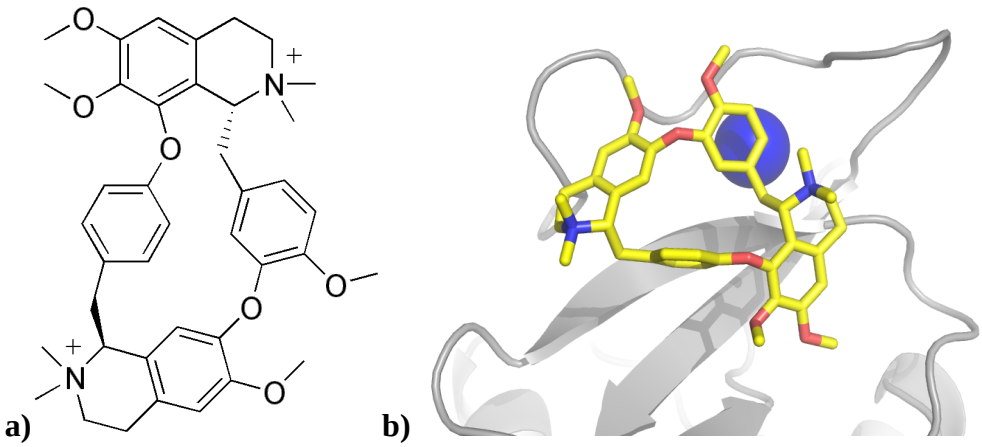

Figura 52: a) Estrutura 2D do Metocurine. **b)** Interação entre MR e Metocurine. O MR está mostrado em *cartoon* cinza. O ligante está mostrado em *sticks* amarelos. O Ca^{2+} está mostrado como esfera azul. Não há ligações de hidrogênio entre o receptor e o ligante.

4.6. Refinamento de poses selecionadas do *virtual screening*

Por fim, foram feitas simulações de dinâmica molecular dos complexos do DC-SIGN e MR com as moléculas selecionadas a partir do *virtual screening* de acordo com diferentes métricas. O principal objetivo dessas simulações é refinar os resultados ao permitirmos que a estrutura tanto do ligante quanto do receptor se adaptem à presença um do outro, melhorando a descrição dos mecanismos moleculares de interação esperados. Finalmente, as trajetórias das simulações de dinâmica molecular obtidas foram usadas para estimar a afinidade relativa de cada ligante pelo receptor através do método MM/PBSA.

Para fins de comparação, o mesmo procedimento foi seguido para estimar a diferença de energia livre relativa dos fragmentos Man_5 e Man_9 pelos receptores. Eles representam os substratos naturais encontrados no vírus da Dengue.

4.6.1. DC-SIGN

Os valores de afinidade relativa obtidos para o DC-SIGN pelos seus substratos naturais mostra que, conforme esperado, a afinidade por Man_9 é maior do que por Man_5. A diferença entre eles se deve principalmente às interações de dispersão. Além disso, é interessante notar que as contribuições eletrostática e de solvatação polar praticamente se anulam, indicando que as interações feitas pelos anéis de carbono do açúcar dominam sua energia livre de interação (Tabela 13).

Já para os ligantes selecionados pela filtragem do *virtual screeening* vemos que o ligante Brown HT, selecionado por ser a molécula com melhor *score* de acordo com o AutoDock Vina, desligou-se da proteína durante as simulações. Curiosamente, a energia de interação estimada com Oxytetracycline foi positiva, indicando que a interação observada nas simulações era apenas metaestável, ou seja, caso a simulação de dinâmica molecular fosse mais longa provavelmente veríamos dissociação do complexo receptor–ligante. Já o ligante Paclitaxel, tem energia de interação baixa aparentemente devido a pouca interação eletrostática com o receptor, não sendo capaz de compensar a penalidade de solvatação polar. Já o ligante Iohexol apresentou o comportamento mais interessante. A afinidade do DC-SIGN por ele é tão grande quanto pelo substrato natural Man_5, sendo que o valor do termo eletrostático é o mesmo para ambos os ligantes. No entanto, enquanto para Man_5 a energia livre de solvatação anula a contribuição eletrostática das interações, isso não acontece no Iohexol. Trata-se da molécula com mais alto potencial para interação com DC-SIGN encontrada nessa série de simulações.

Tabela 13: Energia livre de interação entre DC-SIGN e os substratos naturais Man_5 e Man_9 e ligantes selecionados a partir do *virtual screening*. Os termos que compõem o valor de $\Delta G_{MM/PBSA}$ estão mostrados: interações de dispersão, interações eletrostáticas e solvatação polar. A contribuição da solvatação apolar é pequeno perto das demais. Para fins de comparação o valor do *score* gerado pelo AutoDock Vina está mostrado. As diferenças de entropia do soluto não foram calculadas nesse caso. Portanto os resultados devem ser interpretados apenas comparativamente.

Ligante	Energia de interação com DC-SIGN [kcal/mol]				
	Dispersão	Eletrostática	Solv. polar	$\Delta G_{MM/PBSA}$	Score Vina
Man_9	-36 ± 2	-105 ± 5	102 ± 3	-44 ± 3	
Man_5	-28 ± 1	-105 ± 3	104 ± 2	-33 ± 2	
Iohexol	-11 ± 1	-105 ± 4	87 ± 3	-33 ± 1	-5.1
Paclitaxel	-29 ± 0	-11 ± 1	34 ± 2	-8 ± 2	-7.0
Oxytetracycline	-17 ± 1	-45 ± 2	75 ± 2	11 ± 2	-7.5
Brown HT	0	0	0	0	-9.0

4.6.2. MR

No caso do MR a diferença de afinidade pelos substratos naturais Man_5 e Man_9 é mais perceptível (Tabela 14). Da mesma forma que observado para o DC-SIGN, os termos de solvatação polar e energia eletrostática praticamente se anulam, mostrando que a complexação é, novamente, dominada pelas interações de dispersão.

Já para os ligantes encontrados no *virtual screeening* vemos que os ligantes Metocurine, selecionado pelo *score* do *docking*, e Ceftaroline Fogamil, selecionado considerando a área enterrada do resíduo Leu-694, se desligaram da proteína durante as simulações. O ligante Chloramphenicol teve praticamente a mesma afinidade pelo MR quanto o substrato natural Man_5. Finalmente, o ligante Iohexol teve afinidade pelo MR maior do que ambos os substratos naturais, indicando que se trata de uma molécula com alto potencial de interação no sítio de reconhecimento de carboidratos das CTLs.

Tabela 14: Energia livre de interação entre MR e os substratos naturais Man_5 e Man_9 e ligantes selecionados a partir do *virtual screening*. Os termos que compõem o valor de $\Delta G_{MM/PBSA}$ estão mostrados: interações de dispersão, interações eletrostáticas e solvatação polar. A contribuição da solvatação apolar é pequeno perto das demais. Para fins de comparação o valor do *score* gerado pelo AutoDock Vina está mostrado. As diferenças de entropia do soluto não foram calculadas nesse caso. Portanto os resultados devem ser interpretados apenas comparativamente.

Ligante	Energia de interação com MR [kcal/mol]				
	Dispersão	Eletrostática	Solv. polar	$\Delta G_{MM/PBSA}$	*Score* Vina
Iohexol	-28 ± 1	-151 ± 3	117 ± 3	-67 ± 1	-5.2
Man_9	-38 ± 2	-126 ± 5	115 ± 2	-54 ± 2	
Chloramphenicol	-16 ± 1	-60 ± 1	49 ± 1	-30 ± 1	-5.8
Man_5	-24 ± 1	-89 ± 3	88 ± 2	-28 ± 2	
Metocurine	0	0	0	0	-8.5
Ceftaroline Fogamil	0	0	0	0	-6.9

Apesar da pequena quantidade de moléculas selecionadas para refinamento devido a restrições computacionais, o uso das métricas de número de ligações de hidrogênio e quantidade de área do receptor enterrada foi eficiente na seleção de moléculas com bom potencial de interação com os receptores. Em particular, Iohexol foi identificado como um potencial inibidor tanto do DC-SIGN quanto do MR. Esse possível mecanismo de inibição simultâneo se mostra promissor para inibir a infecção pelo vírus da dengue através das CTLs. O fato de que todas as melhores moléculas selecionadas a partir *virtual screening* usando nossas métricas são carboidratos, derivados de carboidratos ou polióis pode representar uma validação do método usado, já que os receptores sabidamente reconhecem carboidratos. No entanto, é necessário avaliar o comportamento dessas métricas em ensaios de *virtual screening* que tem outros tipos de proteína como métrica, pois a seleção dessas moléculas pode ser apenas um artefato devido ao grande número de grupos doadores e aceptores de hidrogênio presentes nelas. Adicionalmente, o aumento do número de moléculas selecionadas para refinamento e em especial o teste experimental dessas moléculas poderão nos dar informações importantes sobre a eficiência das nossas métricas.

5. CONCLUSÕES

Nossa investigação sobre os mecanismos moleculares do reconhecimento das glicosilações do envelope do vírus da Dengue pelas CTLs expressas nas células do sistema imune teve início com o papel do Ca^{2+} em tal interação. Inicialmente verificamos que é possível identificar os sítios de complexação de Ca^{2+} já conhecidos do DC-SIGN através de simulações de dinâmica molecular nas quais os cátions presentes na sua estrutura cristalográfica foram removidos e há $CaCl_2$ mantido em solução. A especificidade e adequação dessa metodologia foi validada pela ausência de identificação de sítios de complexação de Ca^{2+} em simulações envolvendo o CLEC5A, que sabidamente não liga cálcio[92].

A priori lidamos com o fato de que existem resultados conflitantes na literatura a respeito do número de Ca^{2+} que o CTLD-4 do MR complexa. Enquanto alinhamentos de sequência com outras proteínas da família[87,88] e a estrutura cristalográfica do CTLD-2[91] sugerem que o CTLD-4 complexe apenas o Ca^{2+} principal, há evidências indiretas de que ele possa complexar dois Ca^{2+}, sendo um deles no sítio principal e o outro em um sítio auxiliar diferente de todos os outros membros da família[87,88]. Construímos um modelo do complexo do CTLD-4 do MR e a glicosilação *high-mannose* com base na sua estrutura cristalográfica sem ligante e que apresenta o artefato de troca de domínios (PDB ID 1EGG). As interações MR–manose que observamos são totalmente consistentes com dados de ressonância magnética nuclear[88], validando o modelo construído. Além disso, simulações com $CaCl_2$ na solução identificaram apenas o sítio de complexação principal de Ca^{2+} e não foi possível estabilizar um possível sítio auxiliar a partir de dados da literatura[87,88] mesmo usando simulações dirigidas e com restrições de posição. Finalmente, a dinâmica conformacional do CTLD-4 do MR complexando um Ca^{2+} foi muito similar ao DC-SIGN complexando três Ca^{2+}, sugerindo que esse é o estado saturado de Ca^{2+} de ambas as lectinas. Assim, nosso trabalho fornece evidência extra de que o CTLD-4 do MR complexa apenas um Ca^{2+}.

De acordo com nossos resultados, a complexação de Ca^{2+} ao DC-SIGN e MR tem três efeitos: *i*) diminuição da energia potencial do complexo receptor–cátion; *ii*) diminuição local da flexibilidade da proteína nos *loops* que complexam Ca^{2+}; e *iii*) estabilização estrutural do sítio de reconhecimento de carboidratos na sua conformação ativa. Assim, este trabalho fornece detalhes a nível molecular que enriquecem a hipótese já existente na literatura de que a complexação de Ca^{2+} às CTLs atua como um mecanismo liga/desliga[73-75,78,86-88]. Isso é particularmente relevante considerando a atividade endocítica do DC-SIGN e MR e a capacidade de ambos retornarem para a membrana celular após abandonar seus ligantes nas vesículas ácidas, onde a protonação dos resíduos ácidos do sítio principal leva à perda de

afinidade por Ca^{2+} e, por consequência, alteração conformacional que faz perder afinidade pelo ligante. Quando retorna ao ambiente extracelular, elas podem ligar Ca^{2+} novamente e recuperar sua atividade lectínica.

As diversas simulações e análises executados neste trabalho possibilita-nos propor uma nova hipótese para explicar o papel do Ca^{2+} nas CTLs: É aceito atualmente que o Ca^{2+} principal participa diretamente da interação com açúcar e contribui para a estabilidade do complexo ao ser coordenado pelas hidroxilas do carboidrato[73-75,78,86-88]. No entanto, as contribuições para a energia livre de interação que calculamos pelo método MM/PBSA sugerem que a interação direta carboidrato–Ca^{2+} é, na verdade, desfavorável à formação do complexo devido à penalidade determinada pela dessolvatação do Ca^{2+} principal. Propomos aqui que sua presença no sítio principal é importante porque, além de estabilizar o sítio de complexação de carboidratos na sua conformação ativa, ele altera as propriedades físico-químicas desse sítio e faz com que seus resíduos tenham maior afinidade pelo carboidrato. Nossos achados foram reproduzidos para valores de constante dielétrica interna de 1 a 8 para três complexos lectina–carboidrato: DC-SIGN–Man_4, DC-SIGN–Le^x, MR–Man_4, mostrando robusta consistência quantitativa. Além disso, nossa hipótese é suportada por simulações de dinâmica molecular nas quais as moléculas de água são tratadas explicitamente. A análise da densidade local do solvente nessas simulações aponta que a complexação de carboidratos precisa deslocar pelo menos duas moléculas de água que estão firmemente ligadas ao Ca^{2+} por fazerem parte da sua primeira camada de solvatação e, ao mesmo tempo, participarem de redes de ligação de hidrogênio com os resíduos do sítio principal. A penalidade de dessolvatação do Ca^{2+} principal deve, ainda, ser levada em conta para o desenho de potenciais inibidores das CTLs: é desejável desenhar inibidores que não exijam a remoção das moléculas de água de solvatação ou, quando isso não for possível, é preciso garantir que as interações CTL–ligante compensem esse custo de dessolvatação.

Simulações de dinâmica molecular do DC-SIGN e MR em complexo com os fragmentos da glicosilação *high-mannose* Man_5 e Man_9 permitiram a identificação de resíduos específicos que estão envolvidos no reconhecimento desses açúcares através de uma combinação de interações hidrofóbicas e ligações de hidrogênio: Phe-313, Glu-344, Glu-347, Glu-358, Ser-360, Asn-362, Asn-365 para DC-SIGN; e His-692, Leu-694, Glu-725, Tyr-729, Glu-737, Glu-733, Asn-747, Ile-749 e His-753 para MR. Nossas simulações revelaram, ainda, que há muitas ligações de hidrogênio transientes entre as lectinas e os carboidratos, isso é, que se formam e se rompem várias vezes. Isso está de acordo com a ideia geral de que carboidratos fazem várias interações de baixa afinidade intrínseca com seus receptores[165]. Além disso, os resíduos envolvidos no reconhecimento dos fragmentos Man_5 e Man_9 são diferentes para o DC-SIGN, porém são os mesmos para MR. Isso está de acordo com a já

conhecida capacidade que o DC-SIGN tem de induzir padrões de ativação celular diferentes ao reconhecer as glicosilações *high-mannose* ou *complex-type*[50,51]. Nossos resultados indicam que a capacidade que o DC-SIGN tem de distinguir padrões de glicosilação é ainda mais profunda do que as diferenças entre *high-mannose* e *complex-type*, sendo capaz de diferenciar diferentes tamanhos da glicosilação *high-mannose*.

Essa capacidade do DC-SIGN foi explorada mais profundamente através do uso combinado de simulações de dinâmica molecular, MM/PBSA, simulações dirigidas para cálculo de força de ruptura, *Umbrella Sampling* e igualdade de Jarzynski para avaliar a afinidade relativa do receptor por todos os possíveis rotâmeros da glicosilação *high-mannose* $GlcNAc_2Man_5$ completa. Nossos dados fornecem evidências de que o receptor tem a capacidade de diferenciar pelo menos alguns rotâmeros da mesma glicosilação através do reconhecimento através de conjuntos de resíduos diferentes para cada rotâmero. Essa observação sugere que este pode ser a base do mecanismo molecular que explica a capacidade do DC-SIGN induzir respostas celulares específicas que variam entre patógenos[50-52]. Essa capacidade de diferenciar carboidratos minimamente diferentes pode implicar em diferentes sinais transduzidos para a célula, dando maior plasticidade e especificidade à resposta imune. Por outro lado, o mecanismo de reconhecimento de açúcares robusto mostrado pelo MR é consistente com a ativação celular de forma independente do patógeno detectado.

Exploramos, adicionalmente, um aspecto mais fundamental e metodológico ao aplicar e comparar os métodos *Umbrella Sampling* e igualdade de Jarzynski para estimar a diferença de energia livre de interação lectina–carboidrato. Apesar de ter sido possível obter o mesmo resultado através dos dois métodos, o custo computacional da igualdade de Jarznski foi quase o dobro daquele do *Umbrella Sampling*. Contudo, de acordo com nossa experiência, o custo computacional da igualdade de Jarzynski pode ser menor do que o do *Umbrella Sampling* em sistemas que demoram muito tempo pra atingir o equilíbrio.

Por fim, fizemos ensaios de *virtual screening* usando uma biblioteca de moléculas aprovadas para uso como drogas pelo FDA e um conjunto de programas e *scripts* de implementação própria. Nossa implementação teve como objetivo paralelizar a execução do *software* AutoDock Vina e calcular as métricas de número de ligações de hidrogênio por resíduo e área enterrada por resíduo para cada pose. Demonstramos que, através da combinação dessas métricas, é possível usar o conhecimento sobre o mecanismo de reconhecimento de substratos naturais para filtrar os resultados do *virtual screening*. Isso possibilita a seleção de moléculas promissoras que tenham um mecanismo de interação similar aos substratos naturais em termos dos resíduos do receptor envolvidos na interação. É interessante destacar que todas as melhores moléculas encontradas são carboidratos ou seus análogos. A aplicação futura do mesmo método a outros alvos nos dirá se isso é um artefato

devido ao grande número de doadores e receptores de hidrogênio nessas moléculas ou se, de fato, podemos considerar esse resultado como uma validação de nosso protocolo de filtragem do *virtual screening*. Após refinamento com dinâmica molecular e estimativa da afinidade relativa usando MM/PBSA a molécula Iohexol, um agente de contraste para exames de imagem, foi identificada como a mais promissora para uso contra a infecção do vírus da Dengue por ter alto potencial de interação tanto com o DC-SIGN quanto com o MR. Assim, para continuidade desta investigação a seleção de mais moléculas a partir no nosso ensaio de *virtual screening* e o teste experimental da sua capacidade de inibição de infecção do vírus da Dengue poderá contribuir de forma importante para a descoberta de *leads* contra essa doença.

6. REFERÊNCIAS

1. MESSINA, Jane P. et al. The many projected futures of dengue. **Nature Reviews Microbiology**, v. 13, n. 4, p. 230, 2015.

2. SELCK, Frederic W.; ADALJA, Amesh A.; BODDIE, Crystal R. An estimate of the global health care and lost productivity costs of dengue. **Vector-Borne and Zoonotic Diseases**, v. 14, n. 11, p. 824-826, 2014.

3. HORSTICK, Olaf; TOZAN, Yesim; WILDER-SMITH, Annelies. Reviewing dengue: still a neglected tropical disease?. **PLoS neglected tropical diseases**, v. 9, n. 4, p. e0003632, 2015.

4. GUZMAN, Maria G.; HARRIS, Eva. Dengue. **The Lancet**, v. 385, n. 9966, p. 453-465, 2015.

5. ACHEE, Nicole L. et al. A critical assessment of vector control for dengue prevention. **PLoS neglected tropical diseases**, v. 9, n. 5, p. e0003655, 2015.

6. DUONG, Veasna et al. Asymptomatic humans transmit dengue virus to mosquitoes. **Proceedings of the National Academy of Sciences**, v. 112, n. 47, p. 14688-14693, 2015.

7. MINISTÉRIO DA SAÚDE (BR). SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE VIGILÂNCIA DAS DOENÇAS TRANSMISSÍVEIS. Dengue: diagnóstico e manejo clínico: adulto e criança. 2013.

8. ROTHMAN, Alan L. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. **Nature reviews Immunology**, v. 11, n. 8, p. 532, 2011.

9. DEJNIRATTISAI, Wanwisa et al. Cross-reacting antibodies enhance dengue virus infection in humans. **Science**, v. 328, n. 5979, p. 745-748, 2010.

10. GUZMAN, Maria G.; ALVAREZ, Mayling; HALSTEAD, Scott B. Secondary infection as a risk factor for dengue hemorrhagic fever/dengue shock syndrome: an historical perspective and role of antibody-dependent enhancement of infection. **Archives of virology**, v. 158, n. 7, p. 1445-1459, 2013.

11. KWAN, Tsz Ho et al. Assessing the risk of dengue virus transmission in a non-endemic city surrounded by endemic and hyperendemic areas. **International Journal of Infectious Diseases**, v. 55, p. 99-101, 2017.

12. CAPEDING, Maria Rosario et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. **The Lancet**, v. 384, n. 9951, p. 1358-1365, 2014.

13. CONSTENLA, Dagna; CLARK, Samantha. Financing dengue vaccine introduction in the Americas: challenges and opportunities. **Expert review of vaccines**, v. 15, n. 4, p. 547-559, 2016.

14. WORLD HEALTH ORGANIZATION et al. Updated Questions and Answers related to information presented in the Sanofi Pasteur press release on 30 November 2017 with regards to the dengue vaccine Dengvaxia. Disponível em http://www.who.int/immunization/diseases/dengue/q_and_a_dengue_vaccine_dengvaxia/en/

15. STIASNY, Karin; HEINZ, Franz X. Flavivirus membrane fusion. **Journal of general virology**, v. 87, n. 10, p. 2755-2766, 2006.

16. LIM, Siew Pheng et al. Ten years of dengue drug discovery: progress and prospects. **Antiviral research**, v. 100, n. 2, p. 500-519, 2013.

17.PERERA, Rushika; KUHN, Richard J. Structural proteomics of dengue virus. **Current opinion in microbiology**, v. 11, n. 4, p. 369-377, 2008.

18. MODIS, Yorgo et al. A ligand-binding pocket in the dengue virus envelope glycoprotein. **Proceedings of the National Academy of Sciences**, v. 100, n. 12, p. 6986-6991, 2003.

19. MODIS, Yorgo et al. Structure of the dengue virus envelope protein after membrane fusion. **Nature**, v. 427, n. 6972, p. 313, 2004.

20. KUHN, Richard J. et al. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. **Cell**, v. 108, n. 5, p. 717-725, 2002.

21. LI, Long et al. The flavivirus precursor membrane-envelope protein complex: structure and maturation. **Science**, v. 319, n. 5871, p. 1830-1834, 2008.

22. ZHANG, Xinzheng et al. Dengue structure differs at the temperatures of its human and mosquito hosts. **Proceedings of the National Academy of Sciences**, v. 110, n. 17, p. 6795-6799, 2013.

23. FIBRIANSAH, Guntur et al. Structural changes in dengue virus when exposed to a temperature of 37 C. **Journal of virology**, v. 87, n. 13, p. 7585-7592, 2013.

24. LOK, Shee-Mei et al. Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. **Nature Structural and Molecular Biology**, v. 15, n. 3, p. 312, 2008.

25. YAP, Sally SL et al. Dengue Virus Glycosylation: What Do We Know?. **Frontiers in microbiology**, v. 8, p. 1415, 2017.

26.DEJNIRATTISAI, Wanwisa et al. Lectin switching during dengue virus infection. **Journal of Infectious Diseases**, v. 203, n. 12, p. 1775-1783, 2011.

27. BRYANT, Juliet E. et al. Glycosylation of the dengue 2 virus E protein at N67 is critical for virus growth in vitro but not for growth in intrathoracically inoculated Aedes aegypti mosquitoes. **Virology**, v. 366, n. 2, p. 415-423, 2007.

28. MONDOTTE, Juan A. et al. Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation. **Journal of virology**, v. 81, n. 13, p. 7136-7148, 2007.

29. TASSANEETRITHEP, Boonrat et al. DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. **Journal of Experimental Medicine**, v. 197, n. 7, p. 823-829, 2003.

30. MILLER, Joanna L. et al. The mannose receptor mediates dengue virus infection of macrophages. **PLoS pathogens**, v. 4, n. 2, p. e17, 2008.

31. IDRIS, Fakhriedzwan; MUHARRAM, Siti Hanna; DIAH, Suwarni. Glycosylation of dengue virus glycoproteins and their interactions with carbohydrate receptors: possible targets for antiviral therapy. **Archives of virology**, v. 161, n. 7, p. 1751-1760, 2016.

32. STANLEY, Pamela et al. Essentials of glycobiology. **Varki, A**, 2009.

33. KIRSCHNER, Karl N.; WOODS, Robert J. Solvent interactions determine carbohydrate conformation. **Proceedings of the National Academy of Sciences**, v. 98, n. 19, p. 10541-10545, 2001.

34. NISHIDA, Yoshihiro et al. 1H NMR Analyses of Rotameric Distribution of C5-C6 bonds of D-Glucopyranoses in Solution. **Journal of Carbohydrate Chemistry**, v. 7, n. 1, p. 239-250, 1988.

35. TVAROŠKA, Igor et al. Quantum mechanical and NMR spectroscopy studies on the conformations of the hydroxymethyl and methoxymethyl groups in aldohexosides. **Carbohydrate research**, v. 337, n. 4, p. 353-367, 2002.

36. TVAROŠKA, Igor; CARVER, Jeremy P. Ab initio molecular orbital calculation of carbohydrate model compounds. 6. The gauche effect and conformations of the hydroxymethyl and methoxymethyl groups. **The Journal of Physical Chemistry B**, v. 101, n. 15, p. 2992-2999, 1997.

37. MIURA, Nobuaki et al. A theoretical study of α-and β-D-glucopyranose conformations by the density functional theory. **Chemical physics letters**, v. 419, n. 4-6, p. 326-332, 2006.

38. ROCKWELL, Glen D.; GRINDLEY, T. Bruce. Effect of solvation on the rotation of hydroxymethyl groups in carbohydrates. **Journal of the American Chemical Society**, v. 120, n. 42, p. 10953-10963, 1998.

39. GUPTA, Ramneek; BRUNAK, Søren. Prediction of glycosylation across the human proteome and the correlation to protein function. In: **Biocomputing 2002**. 2001. p. 310-322.

40. BALAJI, Petety V.; QASBA, Pradman K.; RAO, V. S. R. Molecular dynamics simulations of high-mannose oligosaccharides. **Glycobiology**, v. 4, n. 4, p. 497-515, 1994.

41. YAMAGUCHI, Takumi et al. Exploration of Conformational Spaces of High Mannose Type Oligosaccharides by an NMR Validated Simulation. **Angewandte Chemie**, v. 126, n. 41, p. 11121-11124, 2014.

42. MAHLA, Ranjeet Singh et al. Sweeten PAMPs: role of sugar complexed PAMPs in innate immunity and vaccine biology. **Frontiers in immunology**, v. 4, p. 248, 2013.

43. ZIPFEL, Cyril. Early molecular events in PAMP-triggered immunity. **Current opinion in plant biology**, v. 12, n. 4, p. 414-420, 2009.

44. GABIUS, Hans-Joachim et al. From lectin structure to functional glycomics: principles of the sugar code. **Trends in biochemical sciences**, v. 36, n. 6, p. 298-313, 2011.

45. AMBROSI, Moira; CAMERON, Neil R.; DAVIS, Benjamin G. Lectins: tools for the molecular understanding of the glycocode. **Organic & biomolecular chemistry**, v. 3, n. 9, p. 1593-1608, 2005.

46. MCGREAL, Eamon P.; MILLER, Joanna L.; GORDON, Siamon. Ligand recognition by antigen-presenting cell C-type lectin receptors. **Current opinion in immunology**, v. 17, n. 1, p. 18-24, 2005.

47. GEIJTENBEEK, Teunis BH; GRINGHUIS, Sonja I. Signalling through C-type lectin receptors: shaping immune responses. **Nature Reviews Immunology**, v. 9, n. 7, p. 465, 2009.

48. OSORIO, Fabiola; E SOUSA, Caetano Reis. Myeloid C-type lectin receptors in pathogen recognition and host defense. **Immunity**, v. 34, n. 5, p. 651-664, 2011.

49. CAMBI, Alessandra; KOOPMAN, Marjolein; FIGDOR, Carl G. How C type lectins detect pathogens. **Cellular microbiology**, v. 7, n. 4, p. 481-488, 2005.

50. GARCIA-VALLEJO, Juan J.; VAN KOOYK, Yvette. The physiological role of DC-SIGN: a tale of mice and men. **Trends in immunology**, v. 34, n. 10, p. 482-486, 2013.

51. ŠVAJGER, Urban et al. C-type lectin DC-SIGN: an adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. **Cellular signalling**, v. 22, n. 10, p. 1397-1405, 2010.

52. VAN KOOYK, Yvette; GEIJTENBEEK, Teunis BH. DC-SIGN: escape mechanism for pathogens. **Nature Reviews Immunology**, v. 3, n. 9, p. 697, 2003.

53. GRINGHUIS, Sonja I. et al. Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. **Nature immunology**, v. 10, n. 10, p. 1081, 2009.

54. GRINGHUIS, Sonja I. et al. Fucose-specific DC-SIGN signalling directs T helper cell type-2 responses via IKKε-and CYLD-dependent Bcl3 activation. **Nature communications**, v. 5, p. 3898, 2014.

55. HSU, Shih-Chang et al. Functional interaction of common allergens and a C-type lectin receptor, dendritic cell-specific ICAM3-grabbing non-integrin (DC-SIGN), on human dendritic cells. **Journal of Biological Chemistry**, v. 285, n. 11, p. 7903-7910, 2010.

56. LOZACH, Pierre-Yves et al. Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals. **Journal of Biological Chemistry**, v. 280, n. 25, p. 23698-23708, 2005.

57. NAVARRO SANCHEZ, Erika et al. Dendritic cell specific ICAM3 grabbing non integrin is essential for the productive infection of human dendritic cells by mosquito cell derived dengue viruses. **EMBO reports**, v. 4, n. 7, p. 723-728, 2003.

58. ALEN, Marijke MF et al. Crucial role of the N-glycans on the viral E-envelope glycoprotein in DC-SIGN-mediated dengue virus infection. **Antiviral research**, v. 96, n. 3, p. 280-287, 2012.

59. ALEN, Marijke MF et al. Antiviral activity of carbohydrate-binding agents and the role of DC-SIGN in dengue virus infection. **Virology**, v. 387, n. 1, p. 67-75, 2009.

60. ANDERLUH, Marko et al. DC-SIGN antagonists, a potential new class of anti-infectives. **Current medicinal chemistry**, v. 19, n. 7, p. 992-1007, 2012.

61. VARGA, Norbert et al. A multivalent inhibitor of the DC-SIGN dependent uptake of HIV-1 and Dengue virus. **Biomaterials**, v. 35, n. 13, p. 4175-4184, 2014.

62. LUCZKOWIAK, Joanna et al. Pseudosaccharide functionalized dendrimers as potent inhibitors of DC-SIGN dependent Ebola pseudotyped viral infection. **Bioconjugate chemistry**, v. 22, n. 7, p. 1354-1365, 2011.

63. GARBER, Kathleen CA et al. A general glycomimetic strategy yields non-carbohydrate inhibitors of DC-SIGN. **Chemical Communications**, v. 46, n. 36, p. 6747-6749, 2010.

64. PROST, Lynne R. et al. Noncarbohydrate glycomimetics and glycoprotein surrogates as DC-SIGN antagonists and agonists. **ACS chemical biology**, v. 7, n. 9, p. 1603-1608, 2012.

65. MANGOLD, Shane L.; PROST, Lynne R.; KIESSLING, Laura L. Quinoxalinone inhibitors of the lectin DC-SIGN. **Chemical science**, v. 3, n. 3, p. 772-777, 2012.

66. BORROK, M. Jack; KIESSLING, Laura L. Non-carbohydrate inhibitors of the lectin DC-SIGN. **Journal of the American Chemical Society**, v. 129, n. 42, p. 12780-12785, 2007.

67. QUAN, D. Yu et al. Autonomous tetramerization domains in the glycan-binding receptors DC-SIGN and DC-SIGNR. **Journal of molecular biology**, v. 387, n. 5, p. 1075-1080, 2009.

68. DOS SANTOS, Ália et al. Oligomerization domains in the glycan binding receptors DC SIGN and DC SIGNR: Sequence variation and stability differences. **Protein Science**, v. 26, n. 2, p. 306-316, 2017.

69. MENON, Sindhu et al. Binding-site geometry and flexibility in DC-SIGN demonstrated with surface force measurements. **Proceedings of the National Academy of Sciences**, v. 106, n. 28, p. 11524-11529, 2009.

70. POKIDYSHEVA, Elena et al. Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. **Cell**, v. 124, n. 3, p. 485-493, 2006.

71. FEINBERG, Hadar et al. Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. **Journal of Biological Chemistry**, v. 280, n. 2, p. 1327-1335, 2005.

72. TABARANI, Georges et al. DC-SIGN neck domain is a pH-sensor controlling oligomerization SAXS and hydrodynamic studies of extracellular domain. **Journal of Biological Chemistry**, v. 284, n. 32, p. 21229-21240, 2009.

73. FEINBERG, Hadar et al. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. **Science**, v. 294, n. 5549, p. 2163-2166, 2001.

74. MITCHELL, Daniel A.; FADDEN, Andrew J.; DRICKAMER, Kurt. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR subunit organization and binding to multivalent ligands. **Journal of Biological Chemistry**, v. 276, n. 31, p. 28939-28945, 2001.

75. GUO, Yuan et al. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. **Nature Structural and Molecular Biology**, v. 11, n. 7, p. 591, 2004.

76. THÉPAUT, Michel et al. Structure of a glycomimetic ligand in the carbohydrate recognition domain of C-type lectin DC-SIGN. Structural requirements for selectivity and ligand design. **Journal of the American Chemical Society**, v. 135, n. 7, p. 2518-2529, 2013.

77. SUTKEVICIUTE, Ieva et al. Unique DC-SIGN clustering activity of a small glycomimetic: a lesson for ligand design. **ACS chemical biology**, v. 9, n. 6, p. 1377-1385, 2014.

78. FEINBERG, Hadar et al. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. **Journal of Biological Chemistry**, v. 282, n. 6, p. 4202-4209, 2007.

79. EAST, Lucy; ISACKE, Clare M. The mannose receptor family. **Biochimica et Biophysica Acta (BBA)-General Subjects**, v. 1572, n. 2-3, p. 364-386, 2002.

80. TAYLOR, Philip R.; GORDON, Siamon; MARTINEZ-POMARES, Luisa. The mannose receptor: linking homeostasis and immunity through sugar recognition. **Trends in immunology**, v. 26, n. 2, p. 104-110, 2005.

81. ENGERING, Anneke J. et al. The mannose receptor functions as a high capacity and broad specificity antigen receptor in human dendritic cells. **European journal of immunology**, v. 27, n. 9, p. 2417-2425, 1997.

82. HE, Li-Zhen et al. Antigenic targeting of the human mannose receptor induces tumor immunity. **The Journal of Immunology**, v. 178, n. 10, p. 6259-6267, 2007.

83. AZAD, Abul K.; RAJARAM, Murugesan VS; SCHLESINGER, Larry S. Exploitation of the macrophage mannose receptor (CD206) in infectious disease diagnostics and therapeutics. **Journal of cytology & molecular biology**, v. 1, n. 1, 2014.

84. AZAD, Abul K.; SCHLESINGER, Larry S. Mannose receptor (CD206)-mediated imaging in sentinel lymph node localization. **Clinical and Translational Imaging**, v. 3, n. 3, p. 237-245, 2015.

85. TAYLOR, Maureen E.; BEZOUSKA, K.; DRICKAMER, K. Contribution to ligand binding by multiple carbohydrate-recognition domains in the macrophage mannose receptor. **Journal of Biological Chemistry**, v. 267, n. 3, p. 1719-1726, 1992.

86. HITCHEN, Paul G.; MULLIN, Nicholas P.; TAYLOR, Maureen E. Orientation of sugars bound to the principal C-type carbohydrate-recognition domain of the macrophage mannose receptor. **Biochemical Journal**, v. 333, n. 3, p. 601-608, 1998.

87. MULLIN, Nicholas Paul; HALL, Kersten T.; TAYLOR, Maureen E. Characterization of ligand binding to a carbohydrate-recognition domain of the macrophage mannose receptor. **Journal of Biological Chemistry**, v. 269, n. 45, p. 28405-28413, 1994.

88. MULLIN, Nicholas P.; HITCHEN, Paul G.; TAYLOR, Maureen E. Mechanism of Ca2+ and monosaccharide binding to a C-type carbohydrate-recognition domain of the macrophage mannose receptor. **Journal of Biological Chemistry**, v. 272, n. 9, p. 5668-5681, 1997.

89. LIU, Yanshun; EISENBERG, David. 3D domain swapping: as domains continue to swap. **Protein science**, v. 11, n. 6, p. 1285-1299, 2002.

90. FEINBERG, Hadar et al. Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor. **Journal of Biological Chemistry**, v. 275, n. 28, p. 21539-21548, 2000.

91. HU, Zhenzheng et al. Structural Insights into the pH-Dependent Conformational Change and Collagen Recognition of the Human Mannose Receptor. **Structure**, v. 26, n. 1, p. 60-71. e3, 2018.

92. CHEUNG, Ricky et al. Activation of MDL-1 (CLEC5A) on immature myeloid cells triggers lethal shock in mice. **The Journal of clinical investigation**, v. 121, n. 11, 2011.

93. CHEN, Szu-Ting et al. CLEC5A is critical for dengue-virus-induced lethal disease. **Nature**, v. 453, n. 7195, p. 672, 2008.

94. WU, Ming-Fang et al. CLEC5A is critical for dengue virus–induced inflammasome activation in human macrophages. **Blood**, v. 121, n. 1, p. 95-106, 2013.

95. HUANG, Ya-Lang et al. CLEC5A is critical for dengue virus-induced osteoclast activation and bone homeostasis. **Journal of Molecular Medicine**, v. 94, n. 9, p. 1025-1037, 2016.

96. TENG, Ooiean et al. CLEC5A-mediated enhancement of the inflammatory response in myeloid cells contributes to influenza virus pathogenicity in vivo. **Journal of virology**, v. 91, n. 1, p. e01813-16, 2017.

97. CHEN, Szu-Ting et al. CLEC5A regulates Japanese encephalitis virus-induced neuroinflammation and lethality. **PLoS pathogens**, v. 8, n. 4, p. e1002655, 2012.

98. TUNG, Yen-Ting et al. Nanostructured electrochemical biosensor for th0065 detection of the weak binding between the dengue virus and the CLEC5A receptor. **Nanomedicine: Nanotechnology, Biology and Medicine**, v. 10, n. 6, p. 1335-1341, 2014.

99. LO, Yen-Lung et al. Dengue virus infection is through a cooperative interaction between a mannose receptor and CLEC5A on macrophage as a multivalent hetero-complex. **PloS one**, v. 11, n. 11, p. e0166474, 2016.

100. WATSON, Aleksandra A. et al. Structural flexibility of the macrophage dengue virus receptor CLEC5A implications for ligand binding and signaling. **Journal of Biological Chemistry**, v. 286, n. 27, p. 24208-24218, 2011.

101. TAYLOR, Richard D.; JEWSBURY, Philip J.; ESSEX, Jonathan W. A review of protein-small molecule docking methods. **Journal of computer-aided molecular design**, v. 16, n. 3, p. 151-166, 2002.

102. WARREN, Gregory L. et al. A critical assessment of docking programs and scoring functions. **Journal of medicinal chemistry**, v. 49, n. 20, p. 5912-5931, 2006.

103. CHENG, Tiejun et al. Structure-based virtual screening for drug discovery: a problem-centric review. **The AAPS journal**, v. 14, n. 1, p. 133-141, 2012.

104. SHOICHET, Brian K. Virtual screening of chemical libraries. **Nature**, v. 432, n. 7019, p. 862, 2004.

105. HERTIG, Samuel; LATORRACA, Naomi R.; DROR, Ron O. Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. **PLoS computational biology**, v. 12, n. 6, p. e1004746, 2016.

106. CHIAPPORI, Federica et al. Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations. **PLoS computational biology**, v. 8, n. 12, p. e1002844, 2012.

107. GAO, Cen; DESAPHY, Jeremy; VIETH, Michal. Are induced fit protein conformational changes caused by ligand binding predictable? A molecular dynamics investigation. **Journal of computational chemistry**, v. 38, n. 15, p. 1229-1237, 2017.

108. FEIG, Michael et al. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. **The Journal of Physical Chemistry B**, v. 121, n. 34, p. 8009-8025, 2017.

109. FREDDOLINO, Peter L. et al. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. **Structure**, v. 14, n. 3, p. 437-449, 2006.

110. MORTIER, Jérémie et al. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. **Drug Discovery Today**, v. 20, n. 6, p. 686-702, 2015.

111. KOLDSØ, Heidi et al. Permeation, Gating, and Modulation of the TRPA1 Channel in Long-Timescale Molecular Dynamics Simulations. **Biophysical Journal**, v. 112, n. 3, p. 466a, 2017.

112. RAVAL, Alpan et al. Refinement of protein structure homology models via long, all atom molecular dynamics simulations. **Proteins: Structure, Function, and Bioinformatics**, v. 80, n. 8, p. 2071-2079, 2012.

113. OKIMOTO, Noriaki et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. **PLoS computational biology**, v. 5, n. 10, p. e1000528, 2009.

114. ADCOCK, Stewart A.; MCCAMMON, J. Andrew. Molecular dynamics: survey of methods for simulating the activity of proteins. **Chemical reviews**, v. 106, n. 5, p. 1589-1615, 2006.

115. SALOMON FERRER, Romelia; CASE, David A.; WALKER, Ross C. An overview of the Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 3, n. 2, p. 198-210, 2013.

116. PEREZ, Alberto et al. Advances in free-energy-based simulations of protein folding and ligand binding. **Current opinion in structural biology**, v. 36, p. 25-31, 2016.

117. GENHEDEN, Samuel; RYDE, Ulf. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. **Expert opinion on drug discovery**, v. 10, n. 5, p. 449-461, 2015.

118. MASSOVA, Irina; KOLLMAN, Peter A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. **Perspectives in drug discovery and design**, v. 18, n. 1, p. 113-135, 2000.

119. HOU, Tingjun et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. **Journal of chemical information and modeling**, v. 51, n. 1, p. 69-82, 2010.

120. KUHN, Bernd et al. Validation and use of the MM-PBSA approach for drug discovery. **Journal of medicinal chemistry**, v. 48, n. 12, p. 4040-4048, 2005.

121. HOU, Tingjun; YU, Ron. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. **Journal of medicinal chemistry**, v. 50, n. 6, p. 1177-1188, 2007.

122. TAM, Nguyen Minh; NGUYEN, Minh Tho; NGO, Son Tung. Evaluation of the absolute affinity of neuraminidase inhibitor using steered molecular dynamics simulations. **Journal of Molecular Graphics and Modelling**, v. 77, p. 137-142, 2017.

123. MAI, Binh Khanh; LI, Mai Suan. Neuraminidase inhibitor R-125489–a promising drug for treating influenza virus: steered molecular dynamics approach. **Biochemical and biophysical research communications**, v. 410, n. 3, p. 688-691, 2011.

124. THAI, Nguyen Quoc et al. Screening potential inhibitors for cancer target LSD1 from natural products by steered molecular dynamics. **Molecular Simulation**, v. 44, n. 4, p. 335-342, 2018.

125. HUMMER, Gerhard; SZABO, Attila. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. **Proceedings of the National Academy of Sciences**, v. 98, n. 7, p. 3658-3661, 2001.

126. PARK, Sanghyun et al. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. **The Journal of chemical physics**, v. 119, n. 6, p. 3559-3566, 2003.

127. JARZYNSKI, Christopher. Nonequilibrium equality for free energy differences. **Physical Review Letters**, v. 78, n. 14, p. 2690, 1997.

128. COLLIN, Delphine et al. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. **Nature**, v. 437, n. 7056, p. 231, 2005.

129. XIONG, Hui et al. Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. **Theoretical Chemistry Accounts**, v. 116, n. 1-3, p. 338-346, 2006.

130. BAŞTUĞ, Turgut et al. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. **The Journal of chemical physics**, v. 128, n. 15, p. 04B614, 2008.

131. SOUAILLE, Marc; ROUX, Benoît. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. **Computer physics communications**, v. 135, n. 1, p. 40-57, 2001.

132. KUMAR, Shankar et al. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. **Journal of computational chemistry**, v. 13, n. 8, p. 1011-1021, 1992.

133. ZHANG, Haiyang et al. Cooperative binding of cyclodextrin dimers to isoflavone analogues elucidated by free energy calculations. **The Journal of Physical Chemistry C**, v. 118, n. 13, p. 7163-7173, 2014.

134. WEIS, William I. et al. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. **Science**, v. 254, n. 5038, p. 1608-1615, 1991.

135. PEARLMAN, David A. et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. **Computer Physics Communications**, v. 91, n. 1-3, p. 1-41, 1995.

136. ROCHA, Gerd B. et al. Rm1: A reparameterization of am1 for h, c, n, o, p, s, f, cl, br, and i. **Journal of computational chemistry**, v. 27, n. 10, p. 1101-1111, 2006.

137. STEWART, James JP. MOPAC: a semiempirical molecular orbital program. **Journal of computer-aided molecular design**, v. 4, n. 1, p. 1-103, 1990.

138. VAN DER SPOEL, David et al. GROMACS: fast, flexible, and free. **Journal of computational chemistry**, v. 26, n. 16, p. 1701-1718, 2005.

139. BERENDSEN, Herman JC; VAN DER SPOEL, David; VAN DRUNEN, Rudi. GROMACS: a message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1-3, p. 43-56, 1995.

140. LINDORFF LARSEN, Kresten et al. Improved side chain torsion potentials for the Amber ff99SB protein force field. **Proteins: Structure, Function, and Bioinformatics**, v. 78, n. 8, p. 1950-1958, 2010.

141. KIRSCHNER, Karl N. et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. **Journal of computational chemistry**, v. 29, n. 4, p. 622-655, 2008.

142. WANG, Junmei et al. Development and testing of a general amber force field. **Journal of computational chemistry**, v. 25, n. 9, p. 1157-1174, 2004.

143. MARK, Pekka; NILSSON, Lennart. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. **The Journal of Physical Chemistry A**, v. 105, n. 43, p. 9954-9960, 2001.

144. HESS, Berk et al. LINCS: a linear constraint solver for molecular simulations. **Journal of computational chemistry**, v. 18, n. 12, p. 1463-1472, 1997.

145. MIYAMOTO, Shuichi; KOLLMAN, Peter A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. **Journal of computational chemistry**, v. 13, n. 8, p. 952-962, 1992.

146. DARDEN, Tom; YORK, Darrin; PEDERSEN, Lee. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. **The Journal of chemical physics**, v. 98, n. 12, p. 10089-10092, 1993.

147. BUSSI, Giovanni; DONADIO, Davide; PARRINELLO, Michele. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, v. 126, n. 1, p. 014101, 2007.

148. BERENDSEN, Herman JC et al. Molecular dynamics with coupling to an external bath. **The Journal of chemical physics**, v. 81, n. 8, p. 3684-3690, 1984.

149. NOSÉ, SH [Ubar] ICHI. A molecular dynamics method for simulations in the canonical ensemble. **Molecular Physics**, v. 100, n. 1, p. 191-198, 2002.

150. HOOVER, William G. Canonical dynamics: equilibrium phase-space distributions. **Physical review A**, v. 31, n. 3, p. 1695, 1985.

151. PARRINELLO, Michele; RAHMAN, Aneesur. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied physics**, v. 52, n. 12, p. 7182-7190, 1981.

152. MILLER III, Bill R. et al. MMPBSA. py: an efficient program for end-state free energy calculations. **Journal of chemical theory and computation**, v. 8, n. 9, p. 3314-3321, 2012.

153. GOHLKE, Holger et al. Rigidity theory-based approximation of vibrational entropy changes upon binding to biomolecules. **Journal of chemical theory and computation**, v. 13, n. 4, p. 1495-1502, 2017.

154. PAN, Dabo et al. Computational study on the drug resistance mechanism of hepatitis C virus NS5B RNA-dependent RNA polymerase mutants to BMS-791325 by molecular dynamics simulation and binding free energy calculations. **Chemometrics and Intelligent Laboratory Systems**, v. 154, p. 185-193, 2016.

155. KUMARI, Rashmi et al. g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. **Journal of chemical information and modeling**, v. 54, n. 7, p. 1951-1962, 2014.

156. TROTT, Oleg; OLSON, Arthur J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of computational chemistry**, v. 31, n. 2, p. 455-461, 2010.

157. IRWIN, John J.; SHOICHET, Brian K. ZINC− a free database of commercially available compounds for virtual screening. **Journal of chemical information and modeling**, v. 45, n. 1, p. 177-182, 2005.

158. HUB, Jochen S.; DE GROOT, Bert L.; VAN DER SPOEL, David. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. **Journal of Chemical Theory and Computation**, v. 6, n. 12, p. 3713-3720, 2010.

159. WURZBURG, Beth A.; TARCHEVSKAYA, Svetlana S.; JARDETZKY, Theodore S. Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. **Structure**, v. 14, n. 6, p. 1049-1058, 2006.

160. HAMELBERG, Donald; MCCAMMON, J. Andrew. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. **Journal of the American Chemical Society**, v. 126, n. 24, p. 7683-7689, 2004.

161. LI, Zheng; LAZARIDIS, Themis. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. **Journal of the American Chemical Society**, v. 125, n. 22, p. 6636-6637, 2003.

162. BARILLARI, Caterina et al. Classification of water molecules in protein binding sites. **Journal of the American Chemical Society**, v. 129, n. 9, p. 2577-2587, 2007.

163. MICHEL, Julien; TIRADO-RIVES, Julian; JORGENSEN, William L. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. **Journal of the American Chemical Society**, v. 131, n. 42, p. 15403-15411, 2009.

164. GRAVES, Alan P. et al. A Perspective on Water Site Prediction Methods for Structure Based Drug Design. **Current topics in medicinal chemistry**, v. 17, n. 23, p. 2599-2616, 2017.

165. COLLINS, Brian E.; PAULSON, James C. Cell surface biology mediated by low affinity multivalent protein–glycan interactions. **Current opinion in chemical biology**, v. 8, n. 6, p. 617-625, 2004.

166. ASHBURN, Ted T.; THOR, Karl B. Drug repositioning: identifying and developing new uses for existing drugs. **Nature reviews Drug discovery**, v. 3, n. 8, p. 673, 2004.

APÊNDICE A – Arquivo mdp para minimização de energia sem *constraints*

```
define          = -DFLEXIBLE
constraints     = none
integrator      = steep
nsteps          = 5000000
nstlist         = 10
rlist           = 1.0
coulombtype     = pme
rcoulomb        = 1.0
vdw-type        = cut-off
rvdw            = 1.0
nstenergy       = 10
emtol           = 0.1
emstep          = 0.01
cutoff-scheme = Verlet
```

APENDICE B – Arquivo mdp para minimização de energia com *constraints*

```
constraints     = all-bonds
integrator      = steep
nsteps          = 5000000
nstlist         = 10
rlist           = 1.0
coulombtype     = pme
rcoulomb        = 1.0
vdw-type        = cut-off
rvdw            = 1.0
nstenergy       = 10
emtol           = 0.1
emstep          = 0.01
cutoff-scheme = Verlet
```

APENDICE C – Arquivo mdp para termalização dos sistemas que contém proteína, carboidrato, Ca^{2+}, água, Na^{+} e Cl^{-}

```
define          = -DPOSRES
refcoord_scaling= all
constraints     = all-bonds
integrator      = md
nsteps          = 500000
dt              = 0.002
nstxout         = 0
nstvout         = 0
nstxout-compressed = 0
nstenergy       = 5000
nstlist         = 10
compressed-x-grps = system
rlist           = 1.0
DispCorr        = EnerPres
cutoff-scheme = Verlet
coulombtype   = pme
rcoulomb        = 1.0
vdw-type        = cut-off
rvdw            = 1.0
Tcoupl          = v-rescale
tc-grps         = Protein_CA_Other SOL_NA_CL
tau-t           = 0.1          0.1
ref-t           = 309          309
energygrps      = Protein_CA Other SOL NA CL
Pcoupl          = Berendsen
pcoupltype      = isotropic
tau_p           = 0.5
compressibility = 4.5e-5
ref_p           = 1.0
gen_vel         = yes
gen_temp        = 309.0
gen_seed        = -1
```

APENDICE D – Arquivo mdp para execução de simulação de dinâmica molecular dos sistemas que contém proteína, carboidrato, Ca^{2+}, água, Na^{+} e Cl^{-}

```
constraints     = all-bonds
integrator      = md
nsteps          = 10000000    ; 20 ns
dt              = 0.002
nstxout         = 0
nstvout         = 0
nstxout-compressed = 5000
nstenergy       = 5000
nstlist         = 10
compressed-x-grps = system
rlist           = 1.0
DispCorr        = EnerPres
coulombtype     = pme
rcoulomb        = 1.0
cutoff-scheme = Verlet
vdw-type        = cut-off
rvdw            = 1.0
Tcoupl          = nose-hoover
tc-grps         = Protein_CA_Other SOL_NA_CL
tau-t           = 0.5       0.5
ref-t           = 309       309
energygrps      = Protein_CA Other SOL NA CL
Pcoupl          = parrinello-rahman
pcoupltype      = isotropic
tau_p           = 2.0
compressibility = 4.5e-5
ref_p           = 1.0
gen_vel         = no
gen_temp        = 309.0
gen_seed        = -1
```

APENDICE E – Arquivo mdp para execução de simulação de dinâmica molecular dirigida dos sistemas que contém proteína, carboidrato, Ca^{2+}, água, Na^{+} e Cl^{-}

```
constraints          = all-bonds
integrator           = md
nsteps               = 2000000      ; 4ns
dt                   = 0.002
nstxout              = 0
nstvout              = 0
nstxout-compressed   = 500
nstenergy            = 500
nstlist              = 10
compressed-x-grps    = system
rlist                = 1.0
DispCorr             = EnerPres
coulombtype          = pme
rcoulomb             = 1.0
cutoff-scheme        = Verlet
vdw-type             = cut-off
rvdw                 = 1.0
Tcoupl               = nose-hoover
tc-grps              = Protein_CA_Other SOL_NA_CL
tau-t                = 0.5        0.5
ref-t                = 309        309
energygrps           = Protein_CA Other SOL NA CL
Pcoupl               = parrinello-rahman
pcoupltype           = isotropic
tau_p                = 2.0
compressibility      = 4.5e-5
ref_p                = 1.0
gen_vel              = no
gen_temp             = 309.0
gen_seed             = -1
pull                 = yes
pull-nstxout         = 50
pull-nstfout         = 50
pull-ngroups         = 2
pull-group1_name     = Protein_CA ; sugar_binding_site
pull-group2_name     = sugar_roh ; man5 sugar_roh
pull-ncoords         = 1
pull-coord1-groups   = 1 2
pull-coord1-type     = umbrella
pull-coord1-start    = yes
pull-coord1-geometry= distance
pull-coord1-dim      = Y Y Y
pull-coord1-rate     = 0.001 ; nm/ps = 1 nm/ns = 10 A/ns ;
pull-coord1-k        = 6024.0 ; harmonic force constant in kJ.mol^-1.nm^-2
```

APENDICE F – Arquivo mdp para execução de simulação de dinâmica molecular de cada uma das janelas do *Umbrella Sampling* para os sistemas que contém proteína, carboidrato, Ca^{2+}, água, Na^{+} e Cl^{-}

```
constraints              = all-bonds
integrator               = md
nsteps                   = 25000000    ; 50 ns
dt                       = 0.002
nstxout                  = 0
nstvout                  = 0
nstxout-compressed       = 5000 ; 10 ps
nstenergy                = 5000          ; 10 ps
nstlist                  = 10
compressed-x-grps        = system
rlist                    = 1.0
DispCorr                 = EnerPres
coulombtype              = pme
rcoulomb                 = 1.0
cutoff-scheme            = Verlet
vdw-type                 = cut-off
rvdw                     = 1.0
Tcoupl                   = nose-hoover
tc-grps                  = Protein_CA_Other SOL_NA_CL
tau-t                    = 0.5       0.5
ref-t                    = 309       309
energygrps               = Protein_CA Other SOL NA CL
Pcoupl                   = parrinello-rahman
pcoupltype               = isotropic
tau_p                    = 2.0
compressibility          = 4.5e-5
ref_p                    = 1.0
gen_vel                  = no
gen_temp                 = 309.0
gen_seed                 = -1
pull                     = yes
pull-nstxout             = 500   ; 1 ps
pull-nstfout             = 500   ; 1 ps
pull-ngroups             = 2
pull-group1_name         = Protein_CA
pull-group2_name         = sugar_roh
pull-ncoords             = 1
pull-coord1-groups       = 1 2
pull-coord1-type         = umbrella
pull-coord1-start        = yes
pull-coord1-geometry= distance
pull-coord1-dim          = Y Y Y
pull-coord1-rate         = 0.000 ; not moving!!
pull-coord1-k            = 1000.0 ; Soft spring! ; harmonic force constant in kJ.mol^-1.nm^-2
```

APENDICE G – Arquivo mdp para execução de simulação de dinâmica molecular dirigida para calcular a força de ruptura em sistemas que contém proteína, carboidrato, Ca^{2+}, água, Na^{+} e Cl^{-}

```
constraints          = all-bonds
integrator           = md
nsteps               = 75000        ; 150ps
dt                   = 0.002
nstxout              = 0
nstvout              = 0
nstxout-compressed   = 500
nstenergy            = 500
nstlist              = 10
compressed-x-grps    = system
rlist                = 1.0
DispCorr             = EnerPres
coulombtype          = pme
rcoulomb             = 1.0
cutoff-scheme        = Verlet
vdw-type             = cut-off
rvdw                 = 1.0
Tcoupl               = nose-hoover
tc-grps              = Protein_CA_Other SOL_NA_CL
tau-t                = 0.5       0.5
ref-t                = 309       309
energygrps           = Protein_CA Other SOL NA CL
Pcoupl               = parrinello-rahman
pcoupltype           = isotropic
tau_p                = 2.0
compressibility      = 4.5e-5
ref_p                = 1.0
gen_vel              = no
gen_temp             = 309.0
gen_seed             = -1
pull                 = yes
pull-nstxout         = 50
pull-nstfout         = 50
pull-ngroups         = 2
pull-group1_name     = Protein_CA ; sugar_binding_site
pull-group2_name     = sugar ; man5 sugar_roh
pull-ncoords         = 1
pull-coord1-groups   = 1 2
pull-coord1-type     = umbrella
pull-coord1-start    = yes
pull-coord1-geometry= distance ; = direction
pull-coord1-dim      = Y Y Y
pull-coord1-rate     = 0.02 ; nm/ps = 2 nm in 100 ps
pull-coord1-k        = 1000 ; harmonic force constant in kJ.mol^-1.nm^-2
```