

UNIVERSIDADE FEDERAL DE PELOTAS
Programa de Pós-Graduação em Biotecnologia



Dissertação

**Caracterização *in silico* das estruturas e
funcionamento de proteínas quanto a sua
glicosilação**

Rafael dos Santos Woloski

Pelotas, 2018

Rafael dos Santos Woloski

**Caracterização *in silico* das estruturas e funcionamento de proteínas
quanto a sua glicosilação**

Dissertação apresentada ao Programa
de Pós-Graduação em Biotecnologia
da Universidade Federal de Pelotas,
como requisito parcial à obtenção do
título de Mestre em Ciência (área do
Conhecimento: Biotecnologia).

Orientador: Dr. Luciano da Silva Pinto
Comissão de Acompanhamento: Dr. Moisés João Zotti
Dra. Vanessa Galli

Pelotas, 2018

Universidade Federal de Pelotas / Sistema de Bibliotecas Catalogação na Publicação

W866c Woloski, Rafael dos Santos

Caracterização in silico das estruturas e funcionamento de proteínas quanto a sua glicosilação / Rafael dos Santos Woloski ; Luciano da Silva Pinto, orientador. — Pelotas, 2018.

71 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2018.

1. Bioinformática. 2. Dinâmica molecular. 3. Docking. 4. Glicosilação. 5. Expressão heteróloga. I. Pinto, Luciano da Silva, orient. II. Título.

CDD : 620.8

Elaborada por Ubirajara Buddin Cruz CRB: 10/901

BANCA EXAMINADORA

Prof. Dr. Luciano da Silva Pinto (UFPEL, CDTEC)

Prof. Dr. Moisés João Zotti (UFPEL, FAEM)

Profª. Dra. Vanessa Galli (UFPEL, CDTEC)

Agradecimentos

Aos pais, professores, colegas, amigos e familiares.

A todos que contribuiram de alguma forma nessa etapa da minha vida, meu profundo agradecimento.

Lista de Figuras

FIGURA 1 – RMSD DAS PROTEÍNAS ECORL (A), ML-I (B), PHA-L (C) E RICINA (D),.....	38
FIGURA 2 – HISTOGRAMA DA DISTRIBUIÇÃO DO RMSD NAS VERSÕES GLICOSILADAS (GP) E NÃO-GLICOSILADAS (P) DA PROTEÍNA RICINA.	40
FIGURA 3 – HISTOGRAMA DA DISTRIBUIÇÃO DO RMSD DA PROTEÍNA ECORL GLICOSILADA (GP) E NÃO-GLICOSILADA (P).....	40
FIGURA 4 – HISTOGRAMA DA DISTRIBUIÇÃO DO RMSD DA PROTEÍNA ML-I EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P) ...	41
FIGURA 5 – HISTOGRAMA DA DISTRIBUIÇÃO DO RMSD DA PROTEÍNA PHA-L EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P) 41	41
FIGURA 6 – ANÁLISE DE RAIO DE GIRO DAS PROTEÍNAS ECORL (A), ML-I (B), PHA-L (C) E RICINA (D).	42
FIGURA 7 – COMPARAÇÃO DO RMSF DAS PROTEÍNAS ECORL (A), RICINA CADEIA B (B), ML-I CADEIA A (C) E ML-I CADEIA B (D).....	43
FIGURA 8 – COMPARAÇÃO DO RMSF DAS PROTEÍNAS PHA-L (A) E RICINA CADEIA A (B).....	43
FIGURA 9 – PROTEÍNAS ECORL (A), ML-I (B), PHA-L (C) E RICINA (D) VISUALIZADAS POR <i>B-FACTOR</i>	45
FIGURA 10 – PRESENÇA DE GLICINAS CAUSANDO UM AUMENTO NA FLUTUAÇÃO DOS RESÍDUOS	46
FIGURA 11 – ÁREA ACESSÍVEL POR SOLVENTE DAS PROTEÍNAS ECORL (A) E RICINA (B).....	47
FIGURA 12 – ÁREA ACESSÍVEL POR SOLVENTE DA PROTEÍNA PHA-L	47
FIGURA 13 – ÁREA ACESSÍVEL POR SOLVENTE DA PROTEÍNA ML-I	48
FIGURA 14 - ANÁLISE DE CORRELAÇÃO CRUZADA DAS PROTEÍNAS: ML-I GLICOSILADA (A) E NÃO-GLICOSILADA (B); RICINA GLICOSILADA (C) E NÃO-GLICOSILADA (D).	49
FIGURA 15 - ANÁLISE DE CORRELAÇÃO CRUZADA DA PROTEÍNA PHA-L EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P)	50
FIGURA 16 - ANÁLISE DE CORRELAÇÃO CRUZADA DA PROTEÍNA ECORL EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P)	50
FIGURA 17 - SOBREPOSIÇÃO DAS PROTEÍNAS ECORL (A), ML-I (B), PHA-L (C) E RÍCINA (D) JUNTAMENTE COM SEUS LIGANTES JÁ DOCADOS	54
FIGURA 18 - LIGAÇÕES RESULTANTES DO DOCKING DA PROTEÍNA ECOR.....	55
FIGURA 19 - LIGAÇÕES RESULTANTES DO DOCKING DA PROTEÍNA ML-I	55

FIGURA 20 - LIGAÇÕES RESULTANTES DO DOCKING DA PROTEINA PHA-L	56
FIGURA 21 - LIGAÇÕES RESULTANTES DO DOCKING DA PROTEÍNA RICINA	56

Lista de Tabelas

TABELA 1 – OS LIGANTES DE CADA PROTEÍNA UTILIZADOS NA ABORDAGEM DE <i>DOCKING</i>	36
TABELA 2 – MÉDIA DO RMSD DE CADA UMA DAS PROTEÍNAS EM SUA VERSÃO GLICOSILADA (GP) E NÃO-GLICOSILADA (P).	39
TABELA 3 – MÉDIA DO RMSF DAS DIFERENTES PROTEÍNAS.	44
TABELA 4 – NÚMEROS DE CLUSTERS FORMADOS POR CADA UMA DAS PROTEÍNAS EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P).....	51
TABELA 5 – NÚMERO DE MEMBROS NOS DOIS PRIMEIROS CLUSTERS DAS PROTEÍNAS EM SUA FORMA GLICOSILADA (GP) E NÃO-GLICOSILADA (P).	52
TABELA 6 - RESULTADO DOS PROGRAMAS DE DOCKING	53
TABELA 7 - NÚMERO DE PONTES DE HIDROGÊNIO E PONTES SALINAS FORMADAS PELO DOCKING MOLECULAR DAS PROTEÍNAS	57

Resumo

WOLOSKI, Rafael dos Santos. **Caracterização *in silico* das estruturas e funcionamento de proteínas quanto a sua glicosilação.** 2018. 71F. Dissertação (Mestrado) – Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas.

O aumento no interesse científico e comercial por proteínas recombinantes fez com que houvesse uma grande demanda pela sua produção. Entretanto, devido à modificações pós-traducionais de certas classes de proteínas, alguns sistemas heterólogos de expressão não são capazes de produzir de forma funcional algumas proteínas de interesse. A glicosilação é uma modificação pós-traducional presente em mais de 50% dos organismos eucarióticos, sendo a correta glicosilação de algumas proteínas necessária para a sua atividade biológica. Entretanto, sistemas de expressão como *Escherichia coli* não realizam a glicosilação de proteínas, fazendo com que essas glicoproteínas, quando obtidas de forma recombinante, não exerçam sua função biológica. O presente trabalho busca analisar, *in silico*, o comportamento dessas glicoproteínas quanto a presença ou ausência do seu glicano através da abordagem de dinâmica molecular, assim como analisar o seu funcionamento através da abordagem de *docking*. Quatro proteínas foram selecionadas para essa análise, todas pertencentes a classe das lectinas. Suas formas com e sem o glicano foram submetidas a 10 nanosegundos de dinâmica molecular utilizando o programa GROMACS, sendo as estruturas posteriormente clusterizadas e submetidas a *docking* com seus respectivos ligantes. A análise estrutural foi feita utilizando os softwares GROMACS, Bio3D e Pymol. Três programas foram utilizados na abordagem de *docking*: Autodock Vina, CLC Drug Discovery e Swiss-Dock. O programa PLIP foi utilizado para analisar as pontes de hidrogênio da interação. Os resultados mostraram que três das quatro proteínas estudadas tiveram uma maior estabilidade estrutural em suas formas glicosiladas. A abordagem de *docking* também mostrou que as mesmas três proteínas possuem uma melhor afinidade com seus ligantes em sua forma glicosilada. A partir desses resultados, pode-se concluir que é possível utilizar as metodologias de dinâmica e *docking* como uma *pipeline* de análise de estruturas glicosiladas. Entretanto, mais estudos são necessários para verificar se esses resultados vão se manter em um maior tempo de simulação.

Palavras-Chave: Bioinformática, Proteômica, Dinâmica Molecular, *Docking*, Glicosilação

Abstract

WOLOSKI, Rafael dos Santos. **In silico structural and functional analysis of protein glycosylation**, 2018. 71F. Dissertação (Mestrado) – Programa de Pós-Graduação em Biotecnologia. Universidade Federal de Pelotas, Pelotas

The increased scientifical and comercial interest in recombinant proteins caused an increase on demand for its production. However, due to the post-translational modifications of certain proteins, some protein expression systems are incapable of producing some proteins correctly. Protein glycosylation is a post-translational modification that occurs in over 50% of eukaryote proteins, and the correct glycosylation is necessary for the biological activity of some proteins. However, some expression systems, like *Escherichia coli*, does not have the machinery necessary for glycosylation, causing some of these glycoproteins, when produced heterologously, to not function properly. This present work performs an *in silico* analysis of glycoprotein structure and activity in the presence and absence of its glycan, through molecular dynamics and docking. Four proteins, all part of the lectin type, were selected for this analysis; A 10 nanosecond molecular dynamics run was performed for each protein with and without its glycan, using the GROMACS software. These structures were then clustered and docked with their respective ligands. The structural analysis was performed using the GROMACS, Bio3D and PyMol softwares. Docking was performed using the Autodock Vina, CLC Drug Discovery and Swiss-Dock softwares. The PLIP web-server was used to analyze the hydrogen bonds of the docked complex. The results showed that three of the four proteins had a greater structural stability in their glycosylated form. The docking results also showed that the same three proteins had a greater docking affinity in their glycosylated form. These results show that it's possible to combine molecular dynamics and docking as an protein glycosylated structure analysis pipeline. However, further studies are needed to verify if these results can be reproduced in longer simulations.

Keywords: Bioinformatics, Proteomics, Molecular Dynamics, Docking, Glycosylation

Sumário

1	Introdução Geral.....	12
2	Revisão bibliográfica	14
	2.1 Efeito da glicosilação em proteínas.....	14
	2.1.1 O-glicosilação	14
	2.1.2 N-Glicosilação.....	15
	2.2 Abordagem computacional no estudo da estrutura de proteínas	17
	2.3 Dinâmica Molecular.....	19
	2.4 Clustering.....	22
	2.4.1 <i>Geometric clustering</i>	23
	2.4.2 Algoritmos Particionais.....	24
	2.4.3 Algoritmos Hierárquicos	24
	2.5 Docking Molecular.....	25
	2.6 Docking com Receptor Flexível	26
	2.7 Lectinas.....	27
3	Objetivos	30
	3.1 Objetivo Geral	30
	3.2 Objetivos Específicos	30
4	Metodologia.....	31
	4.1 Seleção das Glicoproteínas	31
	4.2 Preparação das Estruturas	32
	4.3 Dinâmica Molecular.....	32
	4.4 Análise estrutural.....	34
	4.5 Produção dos Clusters	35
	4.6 <i>Docking</i> Molecular.....	36
	4.7 Análise dos Resultados de <i>Docking</i>	37
5	Resultados	38
	5.1 Análise Estrutural	38
	5.2 Clusters.....	51
	5.3 Docking Molecular.....	52
6	Discussão.....	58

‘	62
7 Conclusão	63
8 Referências	64

1 1 Introdução Geral

2

3 A descoberta, desenvolvimento, produção e aplicação clínica de glicoproteínas
4 recombinantes para administração terapêutica em humanos tem sido, e
5 continua sendo, uma área de intenso esforço científico e médico. Este esforço
6 gerou interesse considerável nas implicações biológicas e terapêuticas das
7 modificações pós-traducionais, particularmente a mais elaborada e sofisticada,
8 a glicosilação de proteínas.

9 Um dos maiores desafios da moderna Biotecnologia é a produção de proteínas
10 heterólogas em larga escala. Os principais problemas estão relacionados ao
11 correto dobramento das proteínas e ao processamento pós-traducional. Neste
12 sentido, a escolha do sistema de expressão heteróloga adequado (ex:
13 bactérias, leveduras, células de mamíferos) contribui sobremaneira para o êxito
14 do trabalho. A expressão mais barata, mais fácil e rápida de proteínas pode ser
15 realizada em *Escherichia coli*. No entanto, esta bactéria não pode expressar
16 proteínas que requerem modificações pós-tradução complexas como a
17 glicosilação. Por sua vez, a falta da glicosilação nas proteínas heterólogas
18 pode interferir na estabilidade da proteína e causar alterações conformacionais
19 indesejáveis. Estas mudanças podem inclusive modificar a capacidade de
20 ligação destas proteínas ao seu ligante preferencial. Porém, nem toda a
21 glicosilação tem a capacidade de estabilizar a proteína, assim como pode ser
22 irrelevante para o seu funcionamento, e o entendimento destes mecanismos
23 pode ajudar na tomada de decisão da escolha do melhor sistema para
24 expressão de determinadas proteínas.

25 O uso de ferramentas computacionais no estudo de estruturas proteicas tem
26 aumentado a medida que aumenta o interesse comercial na produção de
27 proteínas para uso humano. O grande aumento no número de estruturas
28 tridimensionais de proteínas depositados em bancos de dados como o PDB
29 contribuiu para uma melhor compreensão da sua estrutura tridimensional.
30 Abordagens como *docking* molecular também são utilizadas para analisar o
31 potencial de ligação de uma proteína (receptor) a um ligante (ex: substrato, co-
32 fator), sendo tradicionalmente utilizado para classificar, a partir de uma lista de

33 compostos, aqueles com o maior potencial de ligação a um receptor, visando a
34 descoberta de moléculas biologicamente ativas.

35 Uma das necessidades para se realizar o *docking* molecular com a maior
36 precisão possível é possuir a estrutura de um receptor que melhor demonstre o
37 comportamento da proteína *in vivo*. Isso se torna um problema, já que a
38 maioria das estruturas tridimensionais disponíveis foram determinadas por
39 cristalografia por difração de raios-X, fazendo com que as estruturas
40 resultantes sejam rígidas, não conseguindo simular o estado flexível de
41 proteínas em seu estado natural. Este problema é evidente ao tentar
42 demonstrar, por *docking* molecular, as diferenças supracitadas entre estruturas
43 glicosiladas e não-glicosiladas visto que, ao menos que ambas as estruturas
44 glicosiladas e não-glicosiladas estejam disponíveis, adicionar ou remover
45 glicanos da estrutura 3D produzida por cristalografia não afeta o resultado de
46 uma abordagem de *docking*, devido as mudanças que a presença ou ausência
47 do glicano causam não serem reproduzidas na estrutura tridimensional.

48 A dinâmica molecular é uma abordagem *in silico* que consegue simular o
49 movimento de estruturas proteicas em condições similares aquelas
50 encontradas *in vivo*. Essa abordagem permite a análise da movimentação de
51 uma estrutura tridimensional em um intervalo de tempo especificado,
52 geralmente na casa dos nanosegundos. Deste modo, é possível o uso dessa
53 metodologia para gerar receptores flexíveis a partir de uma estrutura
54 tridimensional rígida. Esses receptores podem então ser utilizados em
55 abordagens de *docking* molecular, conhecidas como *docking com receptor*
56 *flexível*, melhorando assim a precisão da técnica.

57 No presente trabalho, uma abordagem foi desenvolvida utilizando as
58 metodologias de dinâmica molecular e *docking* molecular com receptor flexível
59 para diferenciar proteínas glicosiladas e não-glicosiladas a partir de uma
60 estrutura tridimensional rígida, assim como fazer *docking* nestas estruturas
61 diferenciadas, criando uma *pipeline* que pode ser utilizada para discernir se
62 uma glicoproteína a ser expressa por sistema heterólogo precisa ser glicosilada
63 para exercer corretamente a sua ligação a um receptor.

64

65 **2 Revisão bibliográfica**

66

67 **2.1 Efeito da glicosilação em proteínas**

68

69 Glicosilação é uma modificação pós-traducional em que um glicano, que pode
70 ser um carboidrato monossacarídeo, dissacarídeo ou oligossacarídeo, é ligado
71 covalentemente a um grupo receptor de uma proteína, geralmente um
72 nitrogênio ou oxigênio (NOTHAFT; SZYMANSKI, 2010). Glicosilação é a
73 segunda modificação pós-traducional mais encontrada em proteínas, de acordo
74 com análises feitas no banco de dados Swiss-Prot (KHOURY; BALIBAN;
75 FLOUDAS, 2011), e é uma das estratégias mais comuns utilizadas por
76 sistemas biológicos para aumentar a diversidade do proteoma, podendo se
77 diferenciar pela presença, ausência e pelo tipo de glicano acoplado (WALSH;
78 GARNEAU-TSODIKOVA; GATTO, 2005). A glicosilação afeta a função de
79 proteínas tanto a nível celular, como a sinalização intracelular, quanto a nível
80 proteico, como a estabilidade molecular da proteína, e às vezes são essenciais
81 para o seu funcionamento (SOLÁ; RODRÍGUEZ-MARTÍNEZ; GRIEBENOW,
82 2007).

83 A glicosilação de proteínas pode ser amplamente dividida em dois tipos
84 principais: N-glicosilação e O-glicosilação. Apesar de mais tipos de glicosilação
85 existirem, como C-glicosilação, são eventos que ocorrem com menor
86 frequência na natureza, e menos descritos na literatura (GREENE et al., 2015).

87 **2.1.1 O-glicosilação**

88

89 Na O-glicosilação, um glicano é adicionado ao Oxigênio (O) de uma Serina
90 (Ser), Treonina (Thr) ou Tirosina (Tyr). É um processo que ocorre com menor
91 frequência do que a N-glicosilação, e com um mecanismo menos teoricamente
92 estabelecido, além de não ocorrer em uma sequência conservada como a N-
93 glicosilação. Esse processo ocorre no complexo de Golgi, e geralmente se
94 inicia com a adição de um monossacarídeo, que pode ser uma N-
95 acetilglicosamina, manose, fucose, ácido siálico, N-acetilgalactosamina ou

96 galactose a um dos três resíduos por uma glicosiltransferase, seguido da
97 adição de até mais dois outros carboidratos (SOLÁ; GRIEBENOW, 2009).

98 **2.1.2 N-Glicosilação**

99

100 A N-glicosilação é o processo em que um glicano é adicionado ao nitrogênio
101 (N) de uma Asparagina (Asn), geralmente em uma sequência Asparagina-X-
102 Treonina. É a forma mais abundante e estudada de glicosilação. O processo se
103 inicia no retículo endoplasmático com a adição, através de um lipídio conhecido
104 como dolicol e da ação de uma glicosiltransferase, do oligossacarídeo
105 GlcNAc₂Man₉Gluc₃ no resíduo asparagina de uma proteína. Essa glicoproteína
106 vai subsequentemente passar pelo complexo de Golgi, onde diversas
107 glicosidases e glicosiltransferases vão processar o sacarídeo presente na
108 proteína em diferentes tipos de oligossacarídeos, formando a glicoproteína
109 madura (ZHANG; WANG, 2016).

110 Organismos procariotos, em sua maioria, não fazem glicosilação, fazendo com
111 que sistemas de expressão amplamente utilizados na produção de proteínas
112 recombinantes, como *E. coli*, não produzam glicoproteínas. Além disso,
113 proteínas produzidas em sistemas de expressão heterólogos eucariotos podem
114 ter um padrão de glicosilação diferente daqueles produzidos pelo organismo
115 nativo, já que a atividade de glicosidases e glicosiltransferases, ambas
116 importantes nos processos de glicosilação, tem ação variada nos diferentes
117 organismos (MOENS; VANDERLEYDEN, 1997). *Ad exemplum*, algumas
118 leveduras não conseguem converter oligossacarídeos com alto teor de manose
119 em estruturas complexas (GOOCHEE et al., 1991), assim como os sacarídeos
120 de algumas plantas não possuem ácido siálico, comum em glicoproteínas
121 humanas (KAUSHAL; SZUMILO; ELBEIN, 1988). De mesmo modo, proteínas
122 de plantas são N-glicosiladas com oligossacarídeos formados de N-
123 acetilglicosamina e Galactose, similar a mamíferos, entretanto assim como
124 leveduras, não fazem síntese de ácido siálico. Por fim, também se sabe que,
125 em insetos, ocorre uma glicosilação chamada N-paucimanosídica, que podem
126 possuir fucose ou xilose ligados a uma B-manose (SHI; JARVIS, 2007).

127 Em mamíferos, esse processo é célula-específico, podendo variar de acordo
128 com a presença, concentração e atividade das diferentes glicosidases e
129 glicosiltransferases presentes em cada célula, especialmente N-
130 Acetilglicosaminiltransferases e Alfa-1,3-Galactosiltransferases (PAULSON;
131 COLLEY, 1989). Humanos podem possuir três tipos diferentes de N-
132 glicosilação, dependendo dos diferentes carboidratos compondo o
133 oligossacarídeo: Alto conteúdo de manose, contendo diversas manoses ligadas
134 a duas N-acetilglicosaminas (Exemplo: $\text{Man}_{2-6}\text{Man}_3\text{GlcNAc}_2$), misto ou
135 complexos, que podem ter diferentes tipos de sacarídeos na sua estrutura
136 (Exemplo: $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$) e híbrido, que contem uma cadeia de
137 manose e uma de sacarídeos complexos (Exemplo:
138 $\text{Man}_3\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$) (BROOKS, 2004). Além disso, essas estruturas
139 podem ser complexadas a grupos quimicamente carregados, como fosfatos,
140 ácido carboxílico e ácido siáico em sua porção terminal (SOLÁ; GRIEBENOW,
141 2009).

142 A presença desses glicanos afeta diversas propriedades da proteína, como por
143 exemplo a sua termoestabilidade (MIMURA et al., 2000), resistência a ação de
144 proteases (GOETTIG, 2016), solubilidade (MITRA et al., 2006) e atividade
145 antigênica e imunogênica (GAVRILOV et al., 2011). Além disso, a glicosilação
146 também afeta a estrutura terciária da proteína (SHENTAL-BECHOR; LEVY,
147 2008), e também a atividade ligante-receptor, através de efeitos estéricos e
148 alterações no potencial energético (MARGRAF-SCHÖNFELD; BÖHM; WATZL,
149 2011). Estudos feitos com glicoproteínas mostram que o funcionamento correto
150 de algumas proteínas pode depender do seu padrão de glicosilação, como a
151 presença ou ausência de glicanos, assim como a adição do oligossacarídeo
152 correto. Exemplos disso são a atividade das ácido beta-glicosidases humanas,
153 que dependem da N-glicosilação para exercer sua atividade catalítica (GRACE;
154 GRABOWSKI, 1990). Se tratando de aplicações biotecnológicas, alguns dos
155 fármacos disponíveis no mercado necessitam que sua estrutura seja glicosilada
156 para serem efetivos. Um exemplo disso é a atividade *in vivo* dos hormônios
157 gonadotróficos, que demonstram efeitos variados dependendo da sua
158 glicosilação e, além disso, do tipo de glicano conjugado à proteína (LIU, 1992).

159 Nos últimos anos, a utilização de proteínas recombinantes na terapêutica de
160 diferentes enfermidades expandiu substancialmente o campo da farmacologia
161 molecular. Porém, os produtos farmacêuticos de proteínas apresentam uma
162 série de problemas intrínsecos de instabilidade física e química durante a sua
163 produção, purificação, armazenamento e entrega que podem afetar
164 negativamente suas eficiências terapêuticas finais. Desta forma, tem se
165 buscado intensamente novas estratégias, incluindo a engenharia molecular,
166 para melhorar a estabilidade a longo prazo das proteínas durante o emprego
167 farmacêutico. Além disso, na produção heteróloga de proteínas, o
168 conhecimento das implicações desses glicanos na estabilidade dessas
169 macromoléculas pode ajudar na escolha do melhor sistema de expressão.
170 Cabe ressaltar que devido ao efeito bem conhecido que os glicanos têm no
171 aumento da estabilidade geral das glicoproteínas, a manipulação racional dos
172 parâmetros de glicosilação através da glicogenética pode se tornar uma
173 abordagem promissora para melhorar a estabilidade *in vitro* e *in vivo* de
174 proteínas terapêuticas (SOLÁ; GRIEBENOW, 2009)

175 **2.2 Abordagem computacional no estudo da estrutura de proteínas**
176

177 A importância da estrutura tridimensional de uma proteína está intimamente
178 ligada a determinação da sua função. Isso faz com que seja importante
179 caracterizar a estrutura de uma proteína para alavancar pesquisas sobre suas
180 possíveis aplicações (GOVINDARAJAN; GOLDSTEIN, 1998). Desde a primeira
181 descrição de uma estrutura 3D nos anos 50 (KENDREW et al., 1958), o
182 conhecimento de como são construídas as estruturas macromoleculares cada
183 vez aumenta mais. Bancos de dados como o PDB são uma das plataformas
184 *online* em que estão disponíveis um número diverso de estruturas
185 tridimensionais de proteínas, as quais junto com abordagens como *docking*
186 molecular, podem ser utilizadas para diversos estudos, principalmente no
187 *design* de fármacos, e a simples análise da estrutura 3D de uma proteína já
188 pode gerar um número grande de estudos (MILLS; BEUNING; ONDRECHEN,
189 2015).

190 A maioria das proteínas depositadas nos bancos de dados tiveram sua
191 estrutura descrita por cristalografia por difração de raios-X, o que equivale a

192 mais de 100.000 estruturas depositadas. Essa técnica tem como princípio
193 medir o padrão de difração de raios-X através de uma estrutura cristalina, como
194 a intensidade e o ângulo de difração, uma estrutura tridimensional do objeto
195 pode ser obtida (SHI, 2014). Para uma proteína possa ser descrita por esse
196 método, é necessário que ela forme um cristal, que nada mais é do que um
197 arranjo periódico da estrutura da proteína onde ela se encontra compactada e
198 no seu menor estado energético (ILARI; SAVINO, 2008).

199 Um dos problemas da estrutura obtida por cristalografia é a rigidez da proteína.
200 Proteínas, em seu estado nativo e em solução, apresentam estruturas móveis,
201 e não existem em apenas uma conformação, mas em uma gama de diferentes
202 estados, separados por diferenças na energia do sistema. Essas diferenças
203 vão depender de uma série de fatores, como as características do ambiente,
204 assim como a proteína estar acoplada a um ligante ou não (KUMAR et al.,
205 2000). Esse tipo de discrepância entre a estrutura tridimensional nativa e a
206 estrutura obtida por cristalografia pode causar discrepâncias entre a atividade
207 real de uma proteína em seu estado nativo, o seu comportamento no cristal,
208 quando submetidas a análises *down-stream* como *dockingmolecular*.

209 Um método alternativo utilizado para descrever estruturas proteicas é a técnica
210 de espectroscopia por ressonância nuclear magnética (NMR, do inglês *nuclear*
211 *magnetic ressonance*), que usa técnicas de ressonância nuclear para a
212 determinação da estrutura tridimensional da proteína. Uma vantagem que essa
213 técnica tem sobre a difração de raios-X, é a capacidade de analisar a proteína
214 na presença de um solvente, fazendo com que essa estrutura seja mais
215 biologicamente relevante (GARBUZYNSKIY et al., 2005). Ao invés de só uma
216 única estrutura, a espectroscopia tem como resultado um *ensemble* de
217 estruturas sobrepostas, mostrando a movimentação da proteína no solvente.
218 Entretanto, um dos maiores limitantes da técnica de NMR é o tamanho da
219 proteína, geralmente sendo incapaz de processar estruturas com peso maior
220 que 50 kDa(KRISHNAN et al., 2012). Uma abordagem que pode possivelmente
221 solucionar o problema da rigidez de estruturas 3D é a técnica de dinâmica
222 molecular.

223

224 **2.3 Dinâmica Molecular**

225

226 O estudo de estruturas macromoleculares é um ponto chave para a
227 compreensão de sistemas biológicos. Como mencionado, a função biológica de
228 uma proteína está intimamente ligada a sua estrutura tridimensional, mas, indo
229 mais a fundo, pode-se dizer também que a função de uma proteína está
230 intimamente ligada às interações moleculares dos elementos que a compõe, ou
231 seja, da sua dinâmica interna, juntamente com sua interação com o meio
232 (MICHELETTI, 2013). Entretanto, apesar da grande abrangência da utilização
233 dessas estruturas tridimensionais, elas só mostram um lado das interações
234 moleculares que formam a estrutura. A estrutura da proteína cristalizada é
235 rígida, e na sua forma de menor energia. Proteínas, assim como ácidos
236 nucleicos, são estruturas flexíveis por natureza. Proteínas sofrem mudanças
237 em sua conformação estrutural quando estão exercendo sua função. Como
238 regra, qualquer complexo feito por uma proteína implica em uma modificação
239 estrutural (HOSPITAL et al., 2015).

240 Essa movimentação de estruturas proteicas pode ser facilmente visualizada
241 quando temos os dados de um experimento que caracterizou a proteína por
242 NMR (BOTHWELL; GRIFFIN, 2011), que mostra diversos estados e
243 conformações da estrutura. Mudanças na estrutura são comuns em enzimas. A
244 abertura ou o fechamento de um *loop* pode modificar o acesso do solvente ao
245 sítio ativo, o que pode modificar o ambiente químico ao redor do substrato.
246 Algumas características de proteínas também só podem ser compreendidas
247 quando suas propriedades dinâmicas são levadas em consideração. A difusão
248 de pequenos substratos através de enzimas heme-dependentes, por exemplo,
249 requer o surgimento de canais transitórios na estrutura da proteína
250 (JAKOPITSCH et al., 2005). A abordagem tradicional para a compreensão da
251 influência de diferentes conformações na função da proteína é possuir um
252 grande número da mesma estrutura, obtida de cristais em diversos ambientes e
253 complexada com diversas moléculas.

254 Dinâmica molecular de proteínas é uma técnica que foi inicialmente utilizada na
255 década de 70, e era capaz de simular apenas pequenos resíduos (menos de

256 500 átomos) em um intervalo de tempo menor que 10 picosegundos, sem
257 nenhum solvente (MCCAMMON; GELIN; KARPLUS, 1977). Atualmente, a
258 técnica evoluiu de modo a tornar possível a análise de um sistema
259 biologicamente relevante, como proteínas, proteínas de membrana,
260 ribossomos, nucleossomos, em um intervalo de tempo capaz de mostrar
261 movimentos complexos de regiões a serem estudadas, e na presença de um
262 solvente (HOSPITAL et al., 2015).

263 A construção do sistema começa com a seleção de uma estrutura alvo. A
264 estrutura inicial para a simulação pode ser obtida tanto por estruturas
265 experimentais, como as depositadas no PDB, como por estruturas obtidas
266 através de modelagem molecular. A estrutura pode ser representada em
267 diferentes níveis de detalhe, sendo a representação atômica (cada átomo
268 sendo representado adequadamente e sua energia calculada) a que leva aos
269 melhores resultados (CHILDERS et al., 2017). Selecionada a estrutura, é
270 necessário obter as forças que irão agir sob cada um dos átomos do sistema.
271 Essas forças consistem em conjuntos de equações que vão descrever as
272 interações entre os átomos do sistema e, juntamente com seus parâmetros
273 empíricos, são chamados de Campos de Força (KANDT; ASH; PETER
274 TIELEMAN, 2007).

275 Campos de força são definidos como equações complexas que parametrizam
276 certas medidas e constantes e variáveis físicas, como funções periódicas para
277 o cálculo de rotações, lei de Coulomb para forças de Van Der Waals e
278 interações eletrostática, entre outras - para calcular a energia e força envolvida
279 na interação de cada átomo (MACKERELL et al., 1998). Alguns dos campos de
280 força mais utilizados são CHARMM (CHARMM19, CHARMM22, CHARMM27,
281 etc) (MACKERELL et al., 1998), GROMOS (GROMOS96, GROMOS45A3,
282 GROMOS53A5, etc) (OOSTENBRINK et al., 2004), AMBER (AMBER91,
283 AMBER94, AMBER96, etc) (CORNELL et al., 1996) e OPLS (OPLS-AA)
284 (JORGENSEN; MAXWELL; TIRADO-RIVES, 1996). É importante ressaltar que
285 não existe um campo de força que possa se aplicar à todas as situações. A
286 escolha do campo de força depende das necessidades da simulação.

287 A definição da representação do solvente também é chave na construção do
288 sistema, e novamente a melhor forma é a representação explícita e atômica do
289 solvente, na qual ele é calculado de fato e sua presença é computada, o que
290 consegue simular com precisão os mesmos efeitos que solventes reais teriam
291 em uma molécula, embora isso aumente radicalmente o número de átomos a
292 serem simulados. Íons também podem ser adicionados, dependendo dos
293 requisitos do sistema a ser simulado (CHILDERS et al., 2017)

294 Após a parte estática do sistema ter sido construída, é necessário se assegurar
295 que o sistema não possua conflitos estéricos ou posições geométricas
296 proibidas. Para isso, usa-se a metodologia de minimização de Energia. Esse
297 processo consiste em remover o excesso de energia do sistema, que
298 geralmente é resultado do mal posicionamento de átomos no sistema - átomos
299 mais próximos do que deveriam aumentam a força de Van Der Waals
300 repulsiva, o que aumenta a energia do sistema mais do que deveria. A
301 minimização de energia tenta reposicionar os átomos no sistema, fazendo com
302 que a energia tenda a uma mínima local, ou seja, próximo de zero
303 (CHATZIELEFTHERIOU; ADENDORFF; LAGAROS, 2016).

304 Tendo uma estrutura inicial apropriada, com a energia minimizada, o próximo
305 passo é deixar esse sistema em equilíbrio. Equilibrar um sistema nada mais é
306 do que deixar ele na temperatura e/ou pressão estipulados pelo usuário. Para
307 isso, alguns picosegundos de dinâmica são aplicados para deixar o sistema em
308 equilíbrio (GALLO et al., 2009). Há três principais formas de equilibrar o
309 sistema, dependendo da necessidade do usuário:

- 310 • NVT - Número de átomos (N), Volume (V) e Temperatura (T) constantes;
- 311 • NPT - Número de átomos (N), Pressão (P) e Temperatura (T)
312 constantes;
- 313 • NVE - Número de átomos (N), Volume (V) e Energia (E) constantes.

314 Com o sistema construído, minimizado e equilibrado, após as forças que agem
315 em cada átomo serem obtidas, as leis da física newtoniana clássica são
316 aplicadas para calcular a velocidade e obter a posição individual e trajetória de
317 cada átomo. Para evitar instabilidade, o tempo simulado na equação deve ser

318 menor que o movimento mais rápido da estrutura, geralmente 1 a 2
319 femtosegundos, uma medida que é conhecida como *timestep*. Para obter
320 simulações que sejam biologicamente relevantes, a dinâmica calcula milhões
321 de *timesteps*, dependendo do tempo de simulação. A simulação de dinâmica
322 molecular emula então os movimentos físicos dos átomos em uma proteína
323 presente no ambiente desejado. O resultado bruto de uma metodologia de
324 dinâmica molecular é uma sequência de dados que mostra detalhadamente a
325 movimentação dos átomos em função de um intervalo de tempo, chamada
326 Trajetória (SHAO et al., 2007), que é definida pela interação de todos os
327 átomos especificados no sistema. Os potenciais utilizados nas simulações,
328 como temperatura, pressão, solvente, íons, podem ser prontamente
329 modificados pelo usuário de forma a melhor analisar qualquer uma das
330 características do sistema (PATODIA; BAGARIA; CHOPRA, 2014).
331 Abordagens pós-dinâmica também podem ser utilizadas para organizar as
332 trajetórias da proteína em grupos que sigam algum padrão específico.

333 **2.4 Clustering**

334

335 *Clustering* é uma metodologia de aprendizagem não-supervisionada que pode
336 ser utilizada para qualquer conjunto de dados onde um algoritmo possa medir o
337 grau de distância ou similaridade entre pares de elementos de modo a agrupar
338 objetos com características similares em diferentes grupos (*clusters*). Isso faz
339 com que os membros de um grupo sejam similares entre si, e também
340 diferentes dos membros de outros grupos, de acordo com as características e
341 algoritmos utilizados para gerar os *clusters* (DE PARIS et al., 2015).

342 Clusters, em análises de dinâmica molecular, podem ser gerados por
343 algoritmos de *Kinetic clustering* ou *geometric clustering*. O método de *kinetic*
344 *clustering* é a melhor opção para a formação de *clusters*, pois as conformações
345 são agrupadas de acordo com a superfície de energia potencial do sistema,
346 sendo estruturas que representam um mínimo nessa energia chamadas de
347 estados metaestáveis (KELLER; DAURA; VAN GUNSTEREN, 2010a).
348 Entretanto, essa metodologia apresenta diversas limitações no uso em
349 dinâmica molecular, principalmente pela necessidade de convergência
350 estatística entre todos os pares de dados (conformações) presentes no sistema

351 (ROUDI; AURELL; HERTZ, 2009), o que limita essa abordagem à sistemas de
352 pequena escala. Em estudos de Dinâmica Molecular, o método
353 majoritariamente utilizado é *geometric clustering* (FRACCALVIERI et al., 2011).
354 Em cada *cluster*, existe uma conformação que representa a média dos valores
355 utilizados para a formação desse cluster - quer seja o desvio da raiz da média
356 quadrática (RMSD, *root mean square deviation*) RMSD ou outra métrica. Essa
357 coordenada é chamada de centróide, e representa a posição que melhor
358 descreve as conformações presentes no *cluster*.

359 **2.4.1 Geometric clustering**

360

361 Algoritmos de *geometric clustering* tratam as conformações como pontos em
362 um espaço tridimensional e os agrupam de acordo com a distância entre estes
363 pontos, seguindo os padrões da métrica utilizada, maximizando a
364 homogeniedade de membros de um mesmo cluster assim como maximizando
365 a heterogenicidade entre *clusters* (SHENKIN; MCDONALD, 1994). Nesse caso
366 os *clusters* são interpretados como sendo estruturas em estados metaestáveis,
367 baseado na teoria de que grandes mudanças conformacionais não ocorrem
368 devido à restrições na superfície de energia potencial do sistema, enquanto
369 estruturas que possuem uma energia mínima em comum tendem a apresentar
370 conformações estruturais similares (KELLER; DAURA; VAN GUNSTEREN,
371 2010b).

372 Em análises de dinâmica molecular, os algoritmos de *geometric clustering*
373 agrupam conformações estruturais similares de proteínas, obtidas pelo arquivo
374 de trajetória, em *clusters* distintos, os quais mostram as diferentes
375 conformações que a molécula toma. Assim, esses *clusters* são caracterizados
376 entre grupos que possuem mais ou menos membros, sendo que grupos
377 maiores representem conformações que ocorram mais vezes, assim como
378 grupos menores conformações que aconteçam menos. O sistema também
379 segue uma Distribuição de Boltzmann, o que faz com que estados mais
380 energéticos sejam menos representados do que estados com menor energia
381 (SHAO et al., 2007).

382 Na maioria dos casos de análises de dinâmica molecular, as métricas utilizadas
383 para a formação de um cluster são RMSD (TROYER; COHEN, 1995) e DME
384 (TORDA; VAN GUNSTEREN, 1994). Os algoritmos de *Clustering* podem
385 também ser majoritariamente classificados como Particionais ou Hierárquicos

386 **2.4.2 Algoritmos Particionais**
387

388 Nos algoritmos particionais, o usuário define *a priori* o número (K) de clusters
389 (K -grupos) a serem gerados. O algoritmo então seleciona aleatoriamente um
390 número K decentróides, e adiciona conformações similares a esses centróides,
391 formando K -grupos. A cada passo, os centróides desses K -grupos são
392 recalculados, e as conformações são redistribuídas de acordo com a
393 similaridade à esse novo centróide. Esse processo se repete até que o sistema
394 se estabilize, ou seja, até que não ocorram mais mudanças entre membros de
395 diferentes *clusters*. Esse método possui algumas limitações, como o resultado
396 ser dependente do número de *clusters* especificado, além de perder a
397 eficiência quando lida com trajetórias que gera *clusters* com diferenças muito
398 grandes (ESTER et al., 1996).

399 **2.4.3 Algoritmos Hierárquicos**
400

401 Nos algoritmos hierárquicos, os dados são organizados de forma hierárquica
402 em função da proximidade entre indivíduos, formando um dendrograma que
403 divide os dados em subgrupos menores. A construção do dendrograma pode
404 ser por método divisivo (*top-down*) ou aglomerativo (*bottom-up*). Na abordagem
405 *top-down*, o sistema começa tendo todas as conformações como membros de
406 um único *cluster*, e divide progressivamente os membros em cada nodo. No
407 método *bottom-up*, cada conformação começa sendo considerada como um
408 *cluster* individual, e a cada nodo conformações são agrupadas por similaridade
409 (BERKHIN, 2006). Em ambos os métodos, o dendrograma é cortado no nível
410 em que ele atinge um critério pré-estabelecido, que pode ser o número de
411 clusters, uma métrica mínima de distância (como RMSD), entre outros (WOLF;
412 KIRSCHNER, 2013)

413 **2.5 Docking Molecular**

414

415 Docking molecular é um método computacional que é utilizado para avaliar o
416 potencial de ligação de um ligante (uma proteína ou uma molécula) a um
417 receptor (proteína). Esse processo se inicia com a tentativa, pelos algoritmos
418 de *docking*, também conhecidos como algoritmos de busca, de posicionar
419 corretamente um ligante dentro de um sítio ativo do receptor. Dependendo da
420 metodologia, tanto o receptor como o ligante podem ser tratados como rígidos
421 ou flexíveis, sendo os algoritmos que tratam o ligante como flexível muito mais
422 precisos, apesar de necessitar de mais poder de processamento
423 (ROSENFELD; VAJDA; DELISI, 1995). Os algoritmos de busca de *docking*
424 flexível podem ser divididos em três tipos: sistemáticos, estocásticos ou
425 determinísticos (BROOIJMANS; KUNTZ, 2003). Além disso, é necessário
426 avaliar as possíveis ligações que ocorrem nas diferentes posições das
427 moléculas através de um algoritmo de *score* (KITCHEN et al., 2004).

428 Inicialmente, a metodologia de *docking* era utilizado pra estudar complexos
429 proteína-ligante (KUNTZ et al., 1982), mas o sucesso com esse tipo de análise
430 fez com que a metodologia começasse a ser usada para a descoberta de
431 novos fármacos, testando compostos ativos contra um receptor envolvido em
432 algum processo biológico relevante (BROOIJMANS; KUNTZ, 2003). As
433 técnicas de *docking* podem ser divididas em duas, de acordo com o tipo de
434 ligante a ser utilizado. No *docking* proteína-proteína, é necessário ter a
435 estrutura tridimensional das proteínas a serem analisadas, entretanto nem
436 sempre é necessário especificar um sítio ativo para as proteínas (VAKSER,
437 2014). No *docking* proteína-ligante, além de necessária a estrutura 3D da
438 proteína, é também necessária a estrutura 3D da molécula a ser ligada, assim
439 como especificar o sítio em que o algoritmo de *docking* vai tentar fazer a
440 ligação (GRINTER; ZOU, 2014).

441 Um componente essencial das ferramentas de *docking* é o método de *score*,
442 ou seja, a atribuição de um valor numérico referente a algum parâmetro, ao
443 ligante. Mesmo quando uma ligação é corretamente predita, ela não significa
444 nada se o algoritmo de *score* não consegue diferenciar as posições corretas e
445 incorretas dos ligantes. Geralmente, em ligações proteína-ligante o método de

446 score utilizado é o cálculo de energia livre do sistema, entretanto outros
447 métodos podem ser utilizados. Os algoritmos de score podem ser divididos em
448 três grupos: os baseados em campo de força, os baseados em conhecimento e
449 os empíricos (HUANG; GRINTER; ZOU, 2010).

450 Atualmente os programas de *docking* mais utilizados são Autodock 4
451 (COSCONATI et al., 2010), AutodockVina (TROTT; OLSON, 2010), GOLD
452 (JONES et al., 1997), DOCK (ALLEN et al., 2015), GLIDE (FRIESNER et al.,
453 2004) e Swiss-Dock (GROSIDIER; ZOETE; MICHELIN, 2011). Existem
454 diversos trabalhos que avaliam a qualidade de algoritmos de *docking*,
455 entretanto é importante ressaltar que não existe um único programa que seja
456 empiricamente o melhor em todos os parâmetros, o que da mais liberdade para
457 o usuário escolher o programa a ser usado, desde que cumpra com as
458 necessidades do trabalho (FORLI, 2015).

459 **2.6 Docking com Receptor Flexível**

460

461 As primeiras metodologias de *docking* proteína-ligante criadas eram baseadas
462 no modelo chave-fechadura de ligação, proposto por Emil Fischer em 1894. No
463 modelo chave-fechadura, proteínas e ligantes são tratadas como estruturas
464 rígidas, existindo somente em uma única conformação estável, e sua afinidade
465 é definida por um simples cálculo de geometria entre as estruturas, conhecido
466 como *Rigid-Body Docking* (PAGADALA; SYED; TUSZYNSKI, 2017).
467 Entretanto, quando em meio solúvel, proteínas podem assumir diversas
468 conformações, definidas por leis físicas e distribuições estatísticas (TSAI et al.,
469 1999), assim como ligantes também não são estruturas rígidas. As
470 metodologias então evoluíram, buscando tratar receptores e ligantes como
471 moléculas flexíveis, baseado no modelo de encaixe induzido (KOSHLAND,
472 1963).

473 Apesar de vários programas como AutodockVina, GLIDE e GOLD conseguirem
474 tratar ligantes como estruturas flexíveis, conhecido como *flexible docking*,
475 proteínas são estruturas complexas que possuem milhares de átomos. Essa
476 complexidade faz com que seja difícil calcular um resultado de *docking* levando
477 em conta a estrutura da proteína como flexível (COZZINI et al., 2008). Tendo

478 como base a importância da conformação da proteína na metodologia de
479 *docking*, algumas metodologias foram desenvolvidas para se obter receptores
480 flexíveis, de maior destaque sendo a utilização de dinâmica molecular para
481 flexibilidade do receptor (ALONSO; BLIZNYUK; GREADY, 2006).

482 Dinâmica Molecular, como já descrito, é capaz de analisar as diferentes
483 conformações tomadas por uma proteína em um dado intervalo de tempo. O
484 método chamado de *fully flexible receptor* (FFR) Docking utiliza as diversas
485 conformações de uma proteína resultante de uma metodologia de dinâmica
486 molecular como receptor para resolver o problema da flexibilidade do
487 receptor(VAHL QUEVEDO et al., 2014).

488 Entretanto, o uso dessa metodologia é limitado. Devido ao grande número de
489 conformações obtidas através de uma simulação de Dinâmica Molecular, o
490 *FFR docking* de até mesmo um ligante se torna um processo que necessita de
491 uma grande capacidade computacional. Além disso, dependendo do caso
492 certas conformações da proteína podem ser menos ou mais favoráveis ao
493 *docking*, variando com fatores como por exemplo a obstrução do sítio, ou o
494 afastamento de regiões no sítio ativo(ALONSO; BLIZNYUK; GREADY, 2006).

495 Uma das soluções para esse problema é o uso de técnicas de *clustering* no
496 resultado da dinâmica molecular, que torna possível diminuir a redundância de
497 trajetórias que se repetem, assim como consegue organizar as diferentes
498 conformações que uma proteína toma em grupos de conformações similares,
499 criando uma abordagem mais racional(GIOIA et al., 2017). Como descrito,
500 utilizando essa abordagem é possível afirmar que se um receptor pertencente a
501 um cluster interage de uma forma com um ligante, todos os membros do *cluster*
502 vão se portar de forma semelhante (DE PARIS et al., 2015).

503 **2.7 Lectinas**

504

505 Lectinas são proteínas ou glicoproteínas de origem não-imune que se ligam a
506 carboidratos e açúcares de uma forma específica e reversível, ou que
507 precipitam glico-conjugados (HAMID et al., 2013). São importantes para muitos
508 processos biológicos, devido a sua capacidade de reconhecer carboidratos da
509 superfície celular com alta especificidade, além de sua capacidade de

510 reconhecer carboidratos específicos e de se ligarem a eles sem alterar sua
511 estrutura covalente (WEIS; DRICKAMER, 1996). Lectinas de plantas são
512 usadas como modelo para o estudo do reconhecimento de carboidratos por
513 proteínas, pois individualmente elas demonstram alta sensitividade e, em
514 grupo, uma alta diversidade no reconhecimento de estruturas de carboidratos
515 (LANNOO; VAN DAMME, 2014).

516 Algumas das funções bem caracterizadas das lectinas são a de comunicação
517 celular, metástase em câncer, embriogênese, interações patógeno-hospedeiro,
518 e desenvolvimento de tecidos (SHARON, N.; LIS, 1989). São encontradas em
519 diversos organismos e tecidos, como frutas (SINGH; DEVI; NG, 2014), plantas
520 (VAN PARIJS et al., 1991) e animais (XU et al., 2014). Elas também possuem
521 diversas funções e propriedades relevantes para aplicações medicinais como
522 características anti-HIV (AKKOUH et al., 2015), anti-câncer (YAU et al., 2015),
523 e anti-microbiana (DIAS et al., 2015). Tendo em vista essa diversa gama de
524 atividades, assim como o grande número de organismos que possuem essas
525 proteínas, é de bastante interesse, tanto em pesquisa aplicada quanto em
526 produção para fins econômicos, a produção laboratorial destas lectinas, quer
527 sejam recombinantes ou não.

528 Lectinas de plantas geralmente são isoladas de suas fontes naturais, apesar de
529 isso apresentar diversas desvantagens, como a produção de isoformas
530 (OLIVEIRA; TEIXEIRA; DOMINGUES, 2014). A produção recombinante,
531 principalmente em hospedeiros bacterianos, é uma excelente alternativa para
532 superar esse problema, aumentando a disponibilidade de lectinas, fornecer
533 suprimento contínuo e facilitar a purificação de lectinas que possuam atividades
534 relevantes, até mesmo podendo melhorar a função de lectinas já existentes
535 (OLIVEIRA; TEIXEIRA; DOMINGUES, 2013).

536 Nosso grupo de pesquisa tem estudado lectinas e suas aplicações a mais de
537 10 anos, tendo expressado de forma recombinante e analisado a atividade de
538 algumas dessas proteínas(NETO et al., 2011)(KLAFKE et al., 2016). Dentre as
539 lectinas estudadas estão BVL de *Bauhinia variegata*, Bfl de *B. forficata* que
540 foram produzidas de forma recombinante em bactérias e leveduras, assim
541 como duas lectinas sintéticas, construídas com base em lectinas do grupo das

542 Jacalinas. Os resultados da expressão dessas lectinas, mesmo tendo sucesso,
543 ainda está longe do ideal. A baixa produção e atividade reduzida, quando
544 comparada com suas versões nativas, ainda é um desafio a ser vencido. Esta
545 diferença entre as versões nativa e recombinante pode ser reflexo da falta de
546 glicosilação ou glicosilação deficiente promovida pelos sistemas utilizados.
547 Assim, verificou-se a necessidade de entender melhor as implicações da
548 glicosilação na atividade de lectinas e, concomitantemente, em proteínas
549 heterólogas.

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568 **3 Objetivos**

569

570

571 **3.1 Objetivo Geral**

572

573 Utilizar técnicas de dinâmica molecular combinadas com metodologias de
574 *clustering* e *docking* molecular para corretamente prever a atividade da ligação
575 proteína-ligante de uma glicoproteína em sua forma glicosilada e não-
576 glicosilada, assim como entender o efeito da glicosilação na estrutura dessas
577 proteínas.

578

579

580 **3.2 Objetivos Específicos**

581

582 • Desenvolver metodologia, usando abordagem bioinformática, para o
583 estudo dos efeitos de glicosilação em proteínas selecionadas.

584 • Analisar os efeitos da glicosilação na estabilidade da estrutura
585 tridimensional das proteínas usando dinâmica molecular, comparando
586 suas formas glicosiladas e não-glicosiladas.

587 • Analisar as diferenças na interação proteína-ligante de proteínas
588 glicosiladas com e sem a presença do glicano através da abordagem de
589 *docking*.

590 • Criar uma *pipeline* que oriente a escolha do melhor organismo para
591 produção heteróloga de proteínas glicosiladas com base na
592 funcionalidade desta.

593

594

595

596

597 **4 Metodologia**

598

599 **4.1 Seleção das Glicoproteínas**

600

601 O primeiro passo da metodologia foi selecionar as proteínas a serem utilizadas
602 para o estudo da glicosilação em proteínas. Os seguintes critérios foram 603
utilizados na seleção:

604 1. A proteína deve ser uma glicoproteína N-Glicosilada;
605 2. A sua estrutura tridimensional precisa estar depositada já contendo os
606 seus respectivos glicanos
607 3. Serem do grupo das Lectinas

608 Quatro proteínas que cumpriram esses critérios foram selecionadas:

609 Lectina de *Erythrina corallodendron*

610 A Lectina de *Erythrina corallodendron*, também conhecida como EcorL (Código
611 de acesso no PDB: 1AXZ), é uma lectina de planta. A glicosilação ocorre numa
612 asparagina na posição 17 da proteína (ELGAVISH; SHAANAN, 1998).

613 Lectina de *Viscum album*

614 A Lectina I de *Viscum album*, também conhecida como ML-I (Código de acesso
615 no PDB: 2MLL) é uma proteína que possui duas cadeias diferentes 616
(heterodímero). A cadeia A possui função ribossomo-inativante, removendo 617 uma
adenina da posição 4.324 do rRNA 28S (MISHRA et al., 2004). Já a 618 cadeia B
possui atividade de lectina, sendo beta-galactosídeo específica. A 619 glicosilação
ocorre em três diferentes posições na proteína, 107 (Cadeia A), 92 620 (Cadeia B) e
132 (Cadeia B) (KRAUSPENHAAR et al., 1999).

621 Lectina de *Phaseolus vulgaris*

622 A Fitohemaglutinina-L de *Phaseolus vulgaris*, também conhecida como Pha-L 623
(Código de acesso no PDB: 1G8W), é uma proteína extraída de sementes que 624
possui quatro cadeias homodiméricas. Não há função *in vivo* clara dessa 625
proteína, entretanto é predita a sua atividade na defesa contra patógenos,

626 especialmente caruncho-do-feijão. Sua glicosilação é predita nas posições 12 e
627 60 de todos os monômeros, porém sua estrutura cristalizada só consegue 628
demonstrar as glicosilações nas posições 32 (HAMELRYCK et al., 1996)

629 Lectina de *Ricinus communis*

630 A Lectina de *Ricinus communis*, amplamente conhecida como Ricina (Código de
631 acesso no PDB: 2AAI), é um heterodímero. A Cadeia A é uma proteína 632
ribossomo-inativante, que funciona como uma glicosidase, removendo 633
adeninas do rRNA 28S, o que leva a quebra do RNA e a inativação de 634
ribossomos. Já a Cadeia B é onde se encontra a atividade de lectina da 635
proteína. A Ricina é glicosilada nas posições 10 e 236 da Cadeia A, e 95 e 135 636
da Cadeia B. Essa estrutura, porém, não possui a glicosilação da Cadeia A 637
(RUTENBER et al., 1991).

638 **4.2 Preparação das Estruturas**

639

640 Após a seleção das estruturas, as proteínas selecionadas foram preparadas 641
para o estudo. Água, metais e ligantes foram removidos das estruturas que os 642
continham, deixando somente a proteína com seus glicanos. Isso foi feito para 643
que qualquer diferença apresentada pelas estruturas seja devido a presença ou 644
ausência do glicano, além da dificuldade de abordagens de Dinâmica Molecular 645
descrever as cargas de cofatores metálicos (PERARO et al., 2007). Após isso, 646
dois grupos foram formados, um grupo contendo as quatro proteínas em sua 647
forma glicosilada (GP), enquanto o outro grupo continha as mesmas proteínas, 648
mas com todos os glicanos removidos (P).

649 **4.3 Dinâmica Molecular**

650

651 Após as estruturas estarem preparadas, as oito estruturas resultantes (As 652
quatro proteínas escolhidas em sua forma glicosilada e não-glicosilada), foram 653
submetidas a uma etapa de Dinâmica Molecular, afim de diferenciar as 654
estruturas e também analisar diferenças estruturais entre as formas 655
glicosiladas (GP) e não-glicosiladas (P).

656 Devido a dificuldade que campos de força tem em calcular sistemas complexos
657 com proteínas e glicanos, foi utilizada a ferramenta Glycan Reader (JO et al., 658
2011) para gerar a topologia dos proteínas glicosiladas, para ser possível a 659
aplicação de um campo de força. Após isso, o sistema foi construído de forma 660
idêntica para todas as proteínas: Um cubo foi criado com a proteína no seu 661
centro, com seus eixos XYZ a 10 Angstroms de distância. Esse cubo foi então 662
preenchido com água em forma explícita, que se diferencia do modo implícito 663 por
adicionar de fato moléculas de água, ao invés de tratar como uma força 664 potencial
descrevendo sua ação (PAQUET; VIKTOR, 2015), e adicionados 665 150mM de
KCl para o equilíbrio eletroestático do sistema, posicionados pelo 666 método de
Monte Carlo, um algoritmo que usa cadeias de Markov para 667 posicionar íons
no sistema. Todas as interações foram calculadas utilizando o 668 programa
GROMACS 5.0 (ABRAHAM et al., 2015)

669 Existe um certo equilíbrio necessário na construção inicial do sistema. A 670
maioria dos sistemas de Dinâmica Molecular é simulado com algo chamado 671
Condições Periódicas de Contorno (PBC – *Periodic Boundary Conditions*), o 672 que
significa que o espaço interno do sistema é infinito; se, por exemplo, parte 673 da
proteína passa pela parte de baixo do sistema, ela vai reaparecer na parte 674 de
cima, o que pode gerar diversos problemas, como a proteína interagir com 675 regiões
em que não seja possível haver interações em condições normais. 676 Para evitar
isso, o sistema deve ter um tamanho adequado para cobrir a 677 proteína de
forma que esse tipo de movimento não ocorra. Entretanto, quanto 678 maior o
tamanho da caixa, maior o número de átomos do solvente devem ser 679 adicionados
pra preenche-lá, e quanto mais átomos no sistema, maior a 680 complexidade
do sistema, e mais tempo é necessário para simular o sistema 681 (PAQUET;
VIKTOR, 2015).

682 Após a construção inicial do sistema, o campo de força CHARMM36 (HUANG;
683 MACKERELL, 2013) foi utilizado no sistema. Um *time step* de 2 684
femtossegundos foi utilizado em todos os cálculos. O algoritmo *Particle Mesh 685
Ewald* (PME) (DARDEN; YORK; PEDERSEN, 1993) foi utilizado para calcular 686 as
forças eletroestáticas, e um *cutoff* de 10-12 Angstroms foi utilizado na força 687 de
Van Der Waals para limitar a interação entre moléculas distantes, usando 688 uma
função de troca de força (*Force Switch*) para que a energia das interações

689 não chegue abruptamente a 0 quando chegar no ponto de corte, mas sim 690 comece a diminuir gradualmente até chegar a 0 (STEINBACH; BROOKS, 691 1994). O sistema foi então minimizado por 5000 *steps* (Totalizando 0,01 692 nanosegundos) para remover contatos desfavoráveis e eliminar o excesso de 693 energia do sistema, utilizando o método de *steepest descent*, que calcula o 694 possível movimento de um átomo de acordo com o cálculo da energia 695 potencial, e a cada passo iterativamente movimenta esses átomos de forma 696 que o próximo passo só vai ser dado se o passo anterior tenha conseguido 697 diminuir a energia do sistema (ADCOCK; MCCAMMON, 2006).

698 Depois de minimizado, o sistema foi equilibrado por 25000 *steps* (Totalizando 699 0,05 nanosegundos), no método NVT com temperatura constante de 300K, 700 utilizando o algoritmo do termostato de Nosé-Hoover (HOOVER, 1985), 701 recalculando a velocidade dos átomos do sistema de acordo com a sua energia 702 cinética.

703 Após ter sido propriamente minimizado e equilibrado, o sistema foi simulado 704 por 10 nanosegundos no método NPT, com pressão constante de 1 atm, 705 utilizando o algoritmo do barostato de Parrinello-Rahaman (PARRINELLO; 706 RAHMAN, 1981). O algoritmo LINCS (HESS et al., 1997) foi utilizado para 707 restringir o movimento das ligações de hidrogênio, devido a ocorrerem com 708 menor frequência do que o *time step* utilizado. A cada 10 picosegundos de 709 simulação, as coordenadas foram salvas no arquivo de trajetória, resultando ao 710 todo em 1000 frames para cada proteína, que representam a trajetória da 711 proteína nos 10 nanosegundos simulados.

712 **4.4 Análise estrutural**

713

714 Após concluído o processo de dinâmica molecular, os sistemas foram 715 preparados para a análise estrutural. Toda a água foi removida do sistema, 716 assim como os íons que foram adicionados. A proteína também foi ajustada 717 quanto a rotação no seu eixo. Todos os glicanos foram removidos das 718 proteínas glicosiladas (GP) de forma a não interfirir nas análises das estruturas 719 proteicas.

720 Gráficos de *Root Mean Square Deviation* (RMSD) e *Root Mean Square 721 Fluctuation* (RMSF) foram gerados, mostrando o movimento estrutural referente 722 a estrutura inicial e a flutuação de resíduos, respectivamente. As proteínas 723 também foram visualmente inspecionadas, utilizando a técnica de *b-factor*, que 724 posiciona o valor da flutuação dos aminoácidos, obtidos por RMSF, nos 725 aminoácidos do arquivo PDB, permitindo a visualização das flutuações da 726 proteína na sua estrutura tridimensional.

727 Gráficos do Raio de Giro e da Acessibilidade ao Solvente também foram 728 produzidos. O raio de giro é uma medida da compactação da proteína. Já a 729 Área Acessível ao Solvente mede a área de contato entre a molécula e o 730 solvente

731 O pacote Bio3D da linguagem de programação R (GRANT et al., 2006) 732 também foi utilizado para fazer análises de correlação cruzada (CCA, do inglês 733 *Cross Correlation Analysis*) de cada proteína. Essa técnica analisa correlações 734 nos movimentos dos resíduos, e como isso afeta a flutuação e dinâmica do 735 sistema.

736 4.5 Produção dos Clusters

737

738 Para utilizar os dados de dinâmica molecular na metodologia de docking, as 739 trajetórias de cada proteína foram submetidas a um processo de *clustering*, 740 utilizando a ferramenta gmx cluster do programa Gromacs. O algoritmo 741 GROMOS (DAURA et al., 1999) foi utilizado. Esse algoritmo utiliza o RMSD, 742 que também é utilizado como *cutoff* (sendo um valor estabelecido pelo 743 usuário), para agrupar estruturas de acordo com sua similaridade. No caso, 744 estruturas são agrupadas de forma que o RMSD entre elas seja menor do que 745 o estabelecido como *cutoff*. O algoritmo então seleciona a estrutura que possua 746 o maior número de outras estruturas similares, dentro dos parâmetros de 747 RMSD, e define essa estrutura como centroide, formando juntamente com 748 essas estruturas similares, o primeiro cluster. Essas estruturas são removidas, 749 e o algoritmo repete o processo com as estruturas restantes, até todas serem 750 utilizadas.

751 Como o objetivo de aplicar a metodologia de cluster é obter estruturas que 752
 sejam representativas de toda a trajetória da proteína no tempo simulado, e 753
 como foi observado que existem diferenças entre o RMSD de proteínas 754
 glicosiladas e não-glicosiladas, o *cutoff* estabelecido para cada proteína foi a 755
 média do *Pairwise-RMSD* da sua trajetória. Essa técnica se diferencia do 756
 RMSD normal, que compara as estruturas de uma trajetória com uma estrutura 757
 referência (geralmente a estrutura cristalizada). No método *pairwise*, o RMSD é 758
 calculado para todos os pares de estruturas possíveis no sistema. A média 759
 obtida dessa forma é muito mais representativa do que a média por RMSD 760
 utilizando a estrutura cristalizada como referência, pois significa 761
 movimentações da trajetória em relação a ela mesma, ao invés de em relação 762 ao
 cristal.

763 **4.6 Docking Molecular**

764

765 Após as estruturas serem clusterizadas, a estrutura centroide do *cluster* com o 766
 maior número de membros de cada uma das proteínas foi utilizada para ser 767
 submetida a análises de *docking* molecular. Para obter as coordenadas exatas 768
 do sítio de ligação de cada uma das proteínas, o programa COACH (YANG; 769 ROY;
 ZHANG, 2013) foi utilizado. Como um dos objetivos do estudo é saber a 770 diferença
 da capacidade da ligação proteína-ligante entre proteínas glicosiladas 771 e não-
 glicosiladas, os ligantes escolhidos foram aqueles que tem ligação já 772
 estabelecida experimentalmente com a proteína nativa (Tabela 1).

773

774

775 **Tabela 1** – Os ligantes de cada proteína utilizados na abordagem de *docking*

Proteína	Código de Acesso PDB	Ligante
EcorL	1AXZ	Galactose
ML-I	2MLL	Lactose
Pha-L	1G8W	Galactose
Ricina	2AAI	Lactose

776

777 Para obter resultados com maior abrangência e menor taxa de erro, três 778
779 programas de *docking* foram utilizados: Autodock Vina (Trott; Olson, 780
781 2010), CLC DrugDiscovery e SwissDock (Grosdidier; Zoete; Michelin, 782
783 2011).

784

785 **4.7 Análise dos Resultados de *Docking***

786

787 Os resultados de *docking* das proteínas glicosiladas e não-glicosiladas de cada
788 um dos programas foi comparado. Devido as diferenças nos algoritmos de 789
790 score de cada programa de *docking*, os valores numéricos de cada um deles 791
792 tende a ser em escalas diferentes dos outros, fazendo com que um consenso 793
794 numérico que abranja todos os programas não possa ser usado. Ao invés 795 disso,
796 o método utilizado para avaliar os resultados nos diferentes programas 797 foi o
798 consenso geral de concordância dos resultados. Por exemplo, se um 799 programa
799 prediz que a proteína X em sua forma glicosilada tem melhor 800 afinidade na
801 ligação proteína-ligante do que sua forma não-glicosilada, outro 802 programa deve
802 confirmar essa mesma afinidade para que se forme um 803 consenso.

803 Juntamente com a análise dos resultados, as proteínas docadas também foram
804 analisadas pelo programa PLIP (*Protein-Ligand Interaction Profiler*) (Salentin
805 et al., 2015) quanto as suas diferenças estruturais, como o número de pontes
806 de hidrogênio formadas, presença de pontes salinas, número de resíduos
807 interagindo com o ligante, entre outros. A estrutura das proteínas glicosiladas e 808
809 não-glicosiladas também foram sobrepostas para analisar o posicionamento do 810
811 ligante nas diferentes estruturas.

812

813

814

815

816

807 **5 Resultados**

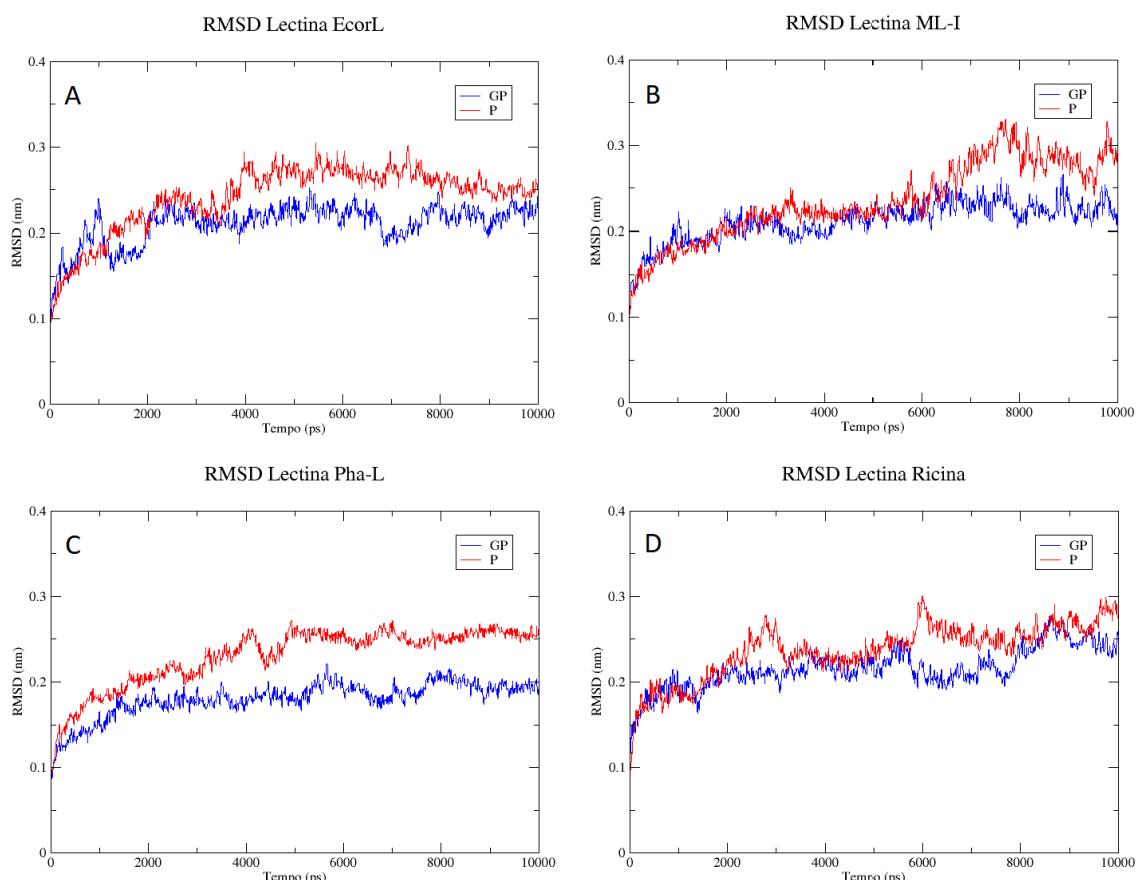
808

809 **5.1 Análise Estrutural**

810

811 Para facilitar as análises, assim como deixar claras quaisquer diferenças, os
 812 gráficos de RMSD das proteínas glicosiladas e não-glicosiladas foram
 813 sobrepostos, assim como os gráficos de RMSF.

814 Nas análises por RMSD, todas as proteínas mostraram a resultados similares,
 815 podendo ser visivelmente observado que o RMSD de proteínas glicosiladas é
 816 menor do que o RMSD de proteínas não-glicosiladas (Figura 1). A Tabela 2 mostra
 817 a diferença entre a média do RMSD dessas proteínas, assim como o seu
 818 DeltaRMSD ($\text{RMSD}_P - \text{RMSD}_{GP}$).



819

820 **Figura 1** – RMSD das proteínas EcorL (A), ML-I (B), Pha-L (C) e Ricina (D), comparando suas
 821 formas glicosiladas (GP, Azul) e não-glicosiladas (P, Vermelho)

822

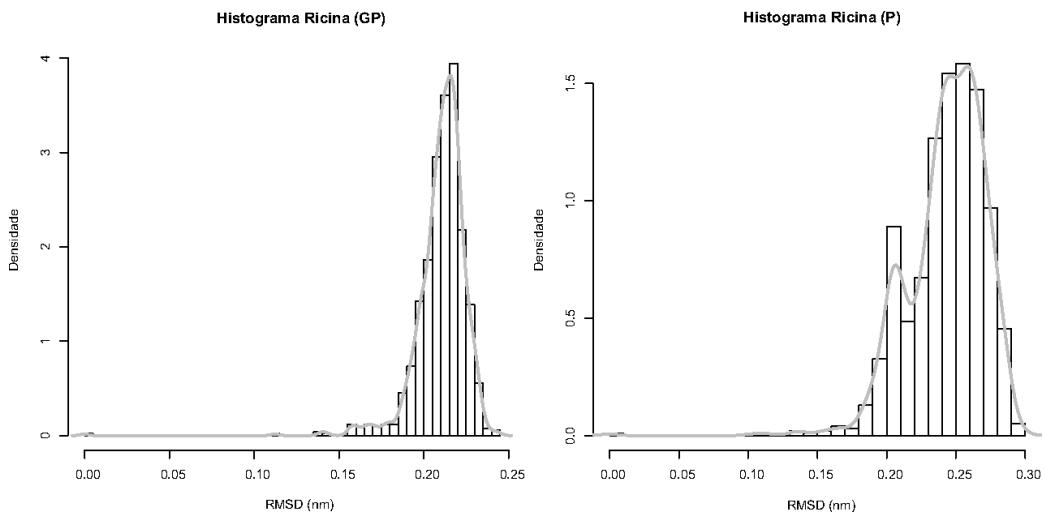
823

824 **Tabela 2** – Média do RMSD de cada uma das proteínas em sua versão glicosilada (GP) e não-
 825 glicosilada (P) com intervalo de confiança, assim como o Delta RMSD ($\text{RMSD}_P - \text{RMSD}_{GP}$).

Proteína	Acesso PDB	N de Glicanos	Forma	Média RMSD	Delta RMSD
EcorL	1AXZ	1	GP	0.209 ± 0.001	
			P	0.241 ± 0.002	0.032
			GP	0.211 ± 0.001	
ML-I	2MLL	3	P	0.235 ± 0.002	0.024
			GP	0.180 ± 0.001	
			P	0.230 ± 0.002	0,05
Pha-L	1G8W	1	GP	0.215 ± 0.001	
			P	0.230 ± 0.002	
			GP	0.237 ± 0.002	0.022
Ricina	2AAI	2	P		

826

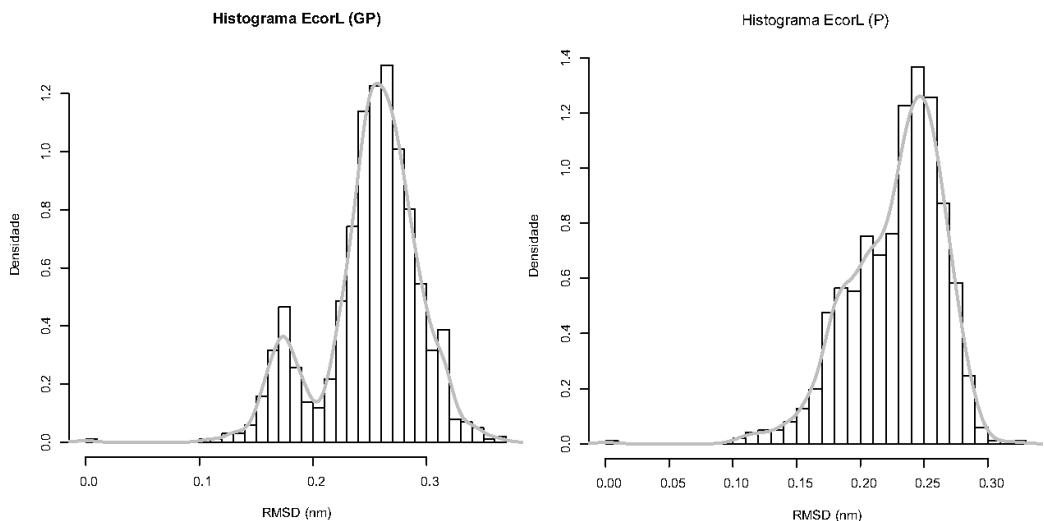
827 A distribuição dos valores de RMSD nas estruturas também foi calculada 828
 utilizando o pacote Bio3D, mas sem sobreposição. Nessa análise, quanto maior 829
 a densidade dos picos, mais conformações da trajetória tiveram o RMSD com o 830
 valor especificado. Em três das proteínas – Ricina (Figura 2), ML-I (Figura 4) e 831
 Pha-L (Figura 5) há picos expressivos e concentrados na densidade do RMSD 832
 das suas formas glicosiladas, o que mostra que essas estruturas tem um 833 RMSD
 mais estável do que sua versão não-glicosilada. Já para a EcorL, o 834 histograma
 teve distribuição similar nas duas formas da proteína (Figura 3).



835

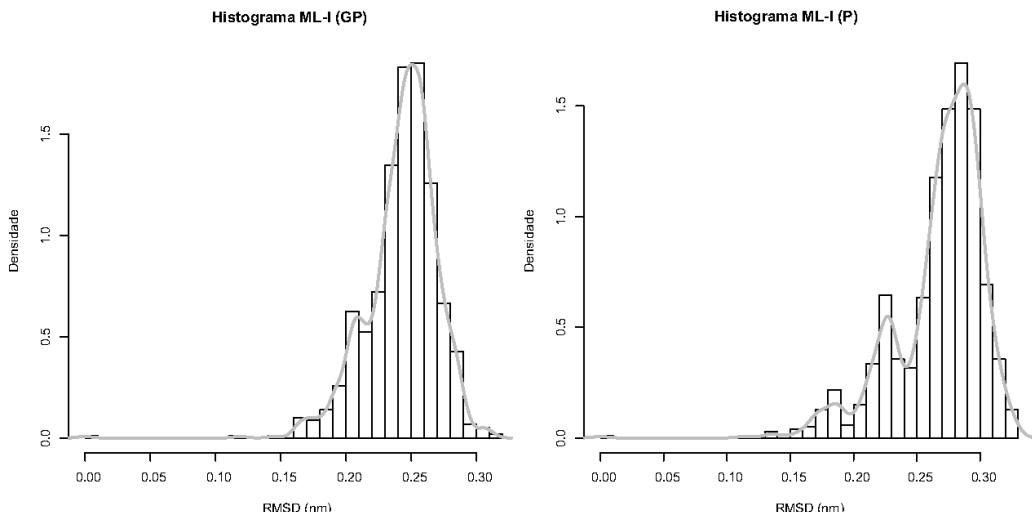
836 **Figura 2** – Histograma da distribuição do RMSD nas versões glicosiladas (GP) e não-
 837 glicosiladas (P) da proteína Ricina. É possível notar um pico expressivo entre os 0.20 e 0.25 da
 838 proteína glicosilada.

839



840

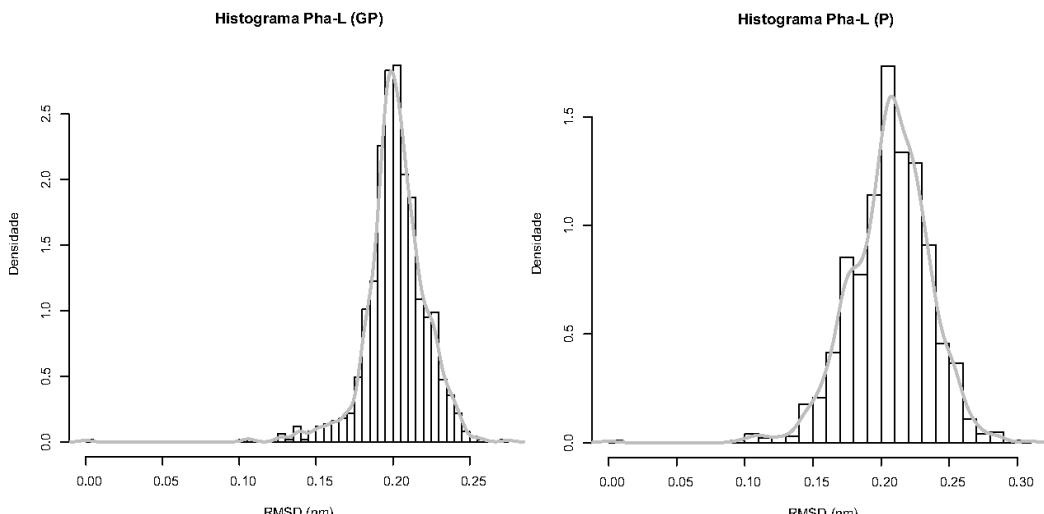
841 **Figura 3** – Histograma da distribuição do RMSD da proteína EcorL glicosilada (GP) e não-
 842 glicosilada (P). A versão glicosilada tem um pequeno pico entre 0.1 e 0.2 nm, ao invés de um
 843 único pico expressivo.



844

845 **Figura 4** – Histograma da distribuição do RMSD da proteína ML-I em sua forma glicosilada
 846 (GP) e não-glicosilada (P).

847



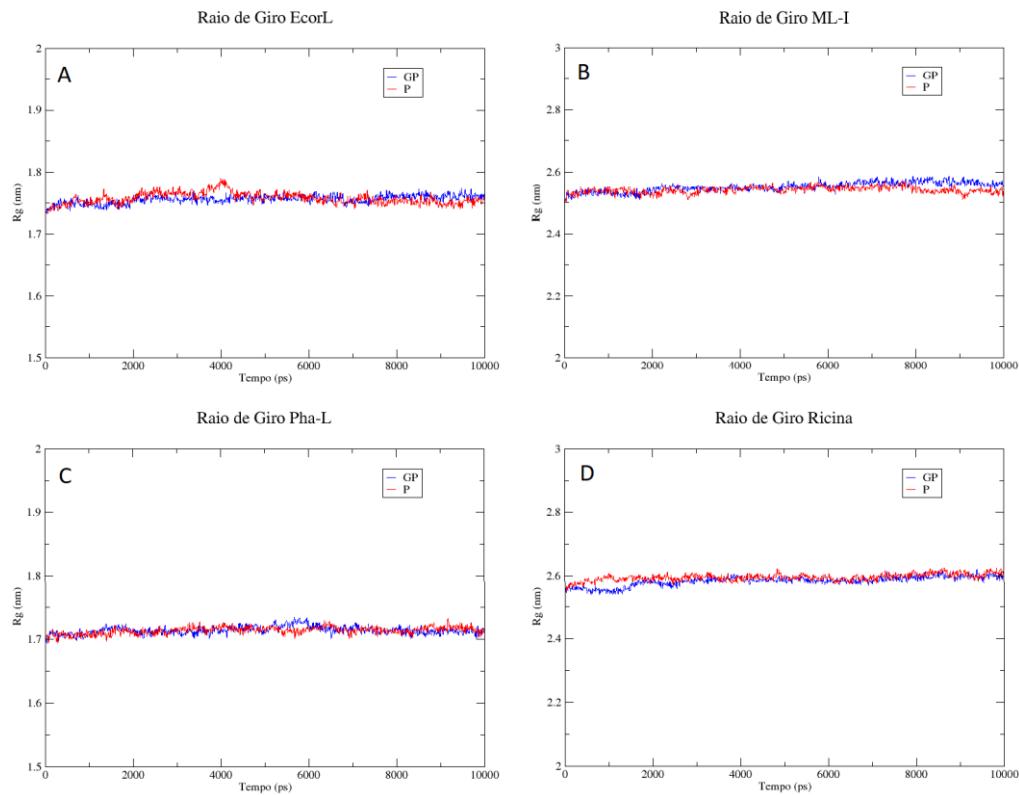
848

849 **Figura 5** – Histograma da distribuição do RMSD da proteína Pha-L em sua forma glicosilada
 850 (GP) e não-glicosilada (P)

851

852

853 Na análise do Raio de Giro das proteínas, não se notou nenhuma diferença 854
 significativa entre as proteínas glicosiladas e não-glicosiladas, o que significa 855 que
 ambas as formas retêm a sua estrutura no mesmo grau de compactação 856 (Figura
 6).



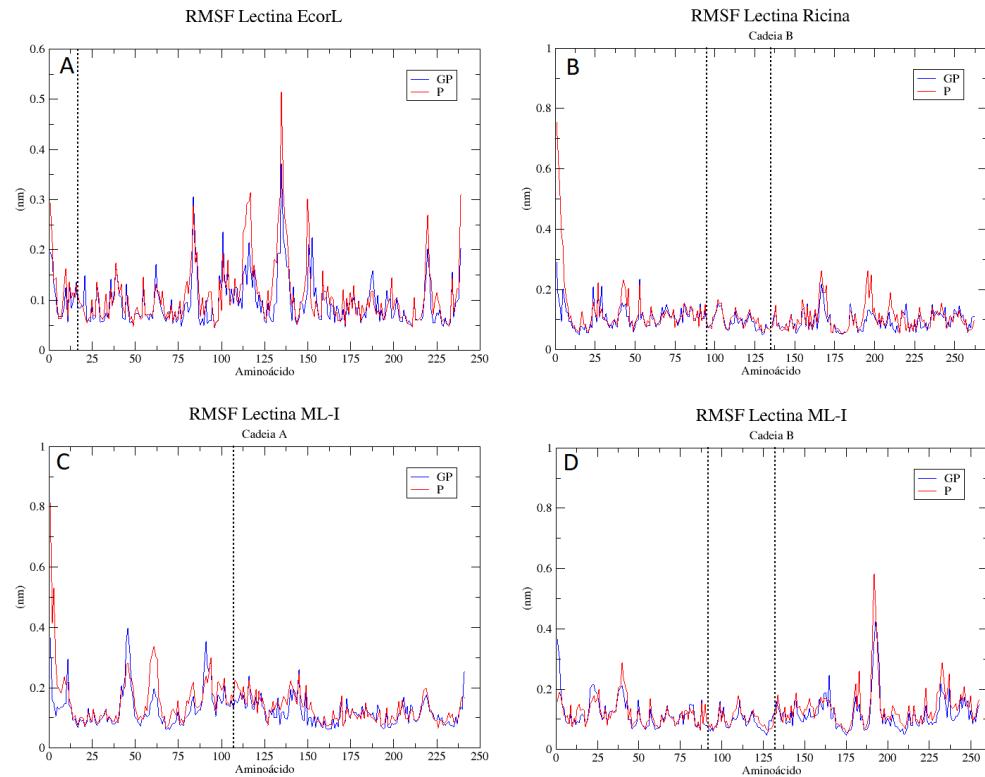
857

858 **Figura 6** – Análise de Raio de Giro das proteínas EcorL (A), ML-I (B), Pha-L (C) e Ricina (D)
 859 em sua forma glicosilada (GP – Azul) e não-glicosilada (P – Vermelho).

860

861

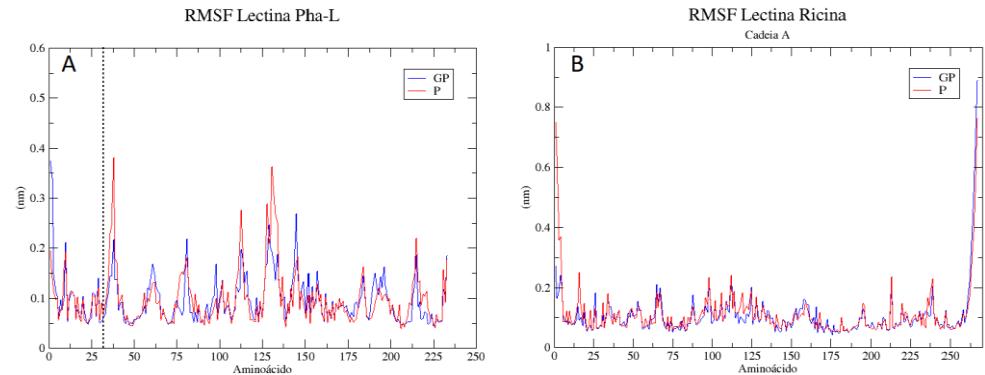
862 Já no cálculo de RMSF, começam a aparecer algumas diferenças entre as 863
 864 proteínas. As proteínas EcorL, ML-I e Ricina cadeia B (Figura 7) possuem uma 864
 865 flutuação menor em sua versão glicosilada. Entretanto, é possível notar tanto 865 na
 866 cadeia A da Ricina quanto na Pha-L que não há muita variação na flutuação 866 entre
 867 as cadeias glicosiladas e não-glicosiladas (Figura 8), como mostra a 867 Tabela 2.
 868 Essa baixa diferenciação da Ricina muito provavelmente se deve à 868 falta de
 869 glicosilação da sua Cadeia A.



869

870 **Figura 7** – Comparação do RMSF das proteínas EcorL (A), Ricina Cadeia B (B), ML-I Cadeia A
 871 (C) e ML-I Cadeia B (D), entre sua forma glicosilada (GP, Azul) e não-glicosilada (P, vermelho).
 872 As linhas serrilhadas verticais representam a localização do sítio de glicosilação.

873



874

875 **Figura 8** – Comparação do RMSF das proteínas Pha-L (A) e Ricina Cadeia A (B), entre suas
 876 formas glicosiladas (GP, Azul) e não-glicosiladas (P, vermelho). As linhas serrilhadas verticais
 877 representam a localização do sítio de glicosilação.

878

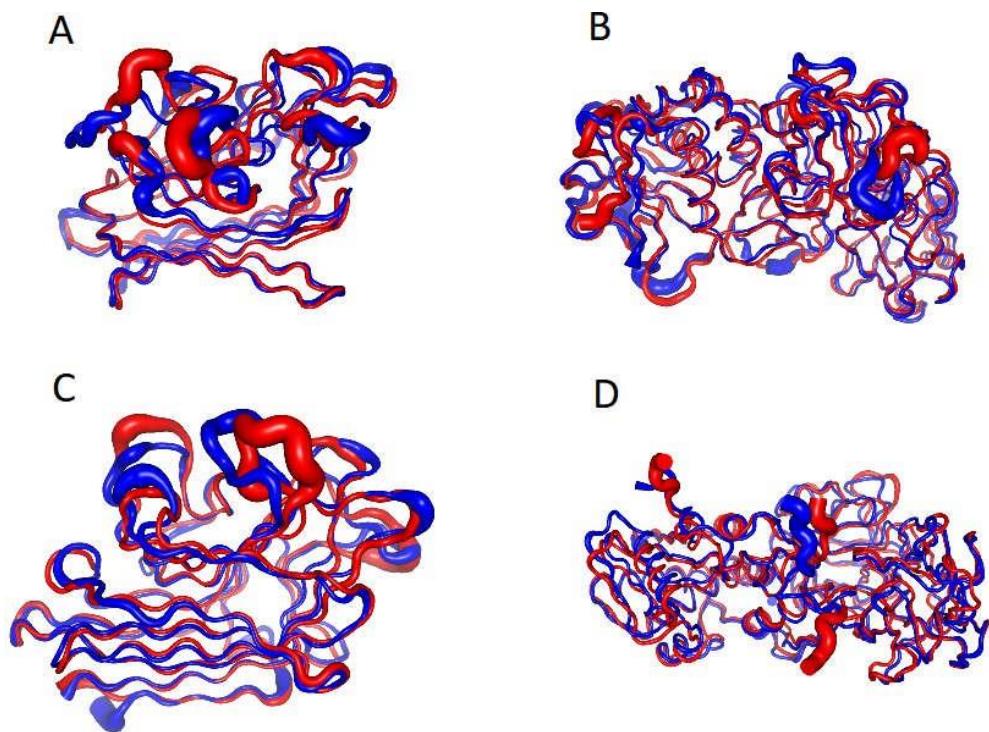
879 **Tabela 3** – Média do RMSF das diferentes proteínas, divididas em suas cadeias, comparando
 880 sua versão glicosilada (GP) e não-glicosilada (P), assim como o número e a localização dos
 881 sítios de glicosilação.

Proteína	PDB	N. Glicanos	Sítio de Glicosilação	Forma	RMSF	DeltaRMSF
				GP	0.096±0.006	
EcorL	1AXZ	1	ASN-17	P	0.110±0.007	0.014
				GP	0.125±0.007	
ML-I A	2MLLa	1	ASN-107	P	0.144±0.009	0,019
				GP	0.115±0.006	
ML-I B	2MLLb	2	ASN-92 ASN-132	P	0.127±0.007	0.012
				GP	0.096±0.006	
Pha-L	1G8W	1	ASN-12	P	0.099±0.007	0.003
				GP	0.106±0.009	
Ricina A	2AAIa	0	-	P	0.112±0.010	0.006
				GP	0.099±0.004	
Ricina B	2AAIb	2	ASN-95 ASN-135	P	0.114±0.008	0.015
				GP		

882

883

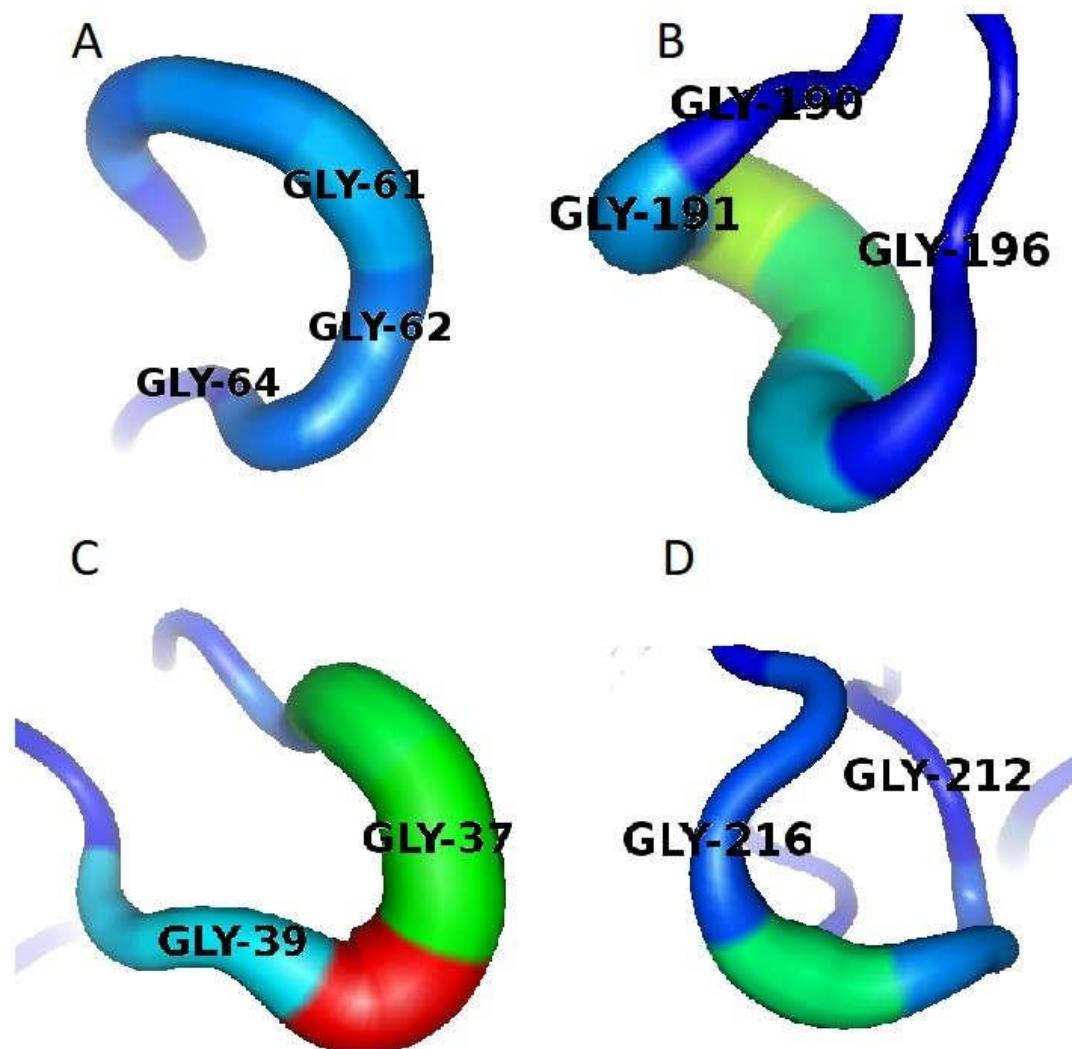
884 As estruturas das proteínas glicosiladas e não-glicosiladas também foram 885
 comparadas por *b-factor*, onde foi possível ver na estrutura 3D a flutuação 886
 mostrada pelo RMSF (Figura 9). Nessa representação, quanto maior a 887
 densidade da estrutura, maior o movimento dela. Os picos do gráfico de RMSF 888
 podem ser facilmente identificados como as estruturas mais densas. Utilizando 889
 esse método, é possível observar que boa parte das estruturas com maior 890
 flutuação nas proteínas ML-I e Pha-L são precedidas pelo aminoácido Glicina, 891
 que se sabe possuir maior flexibilidade (Figura 10) (JACOB; DUCLOHIER; 892
 CAFISO, 1999). A visualização foi feita pelo programa PyMol (DELANO, 2002).



893

894 **Figura 9** – Proteínas EcorL (A), ML-I (B), Pha-L (C) e Ricina (D) visualizadas por *b*-factor em
895 suas formas glicosiladas (Azul) e nã-glicosiladas (Vermelha)

896



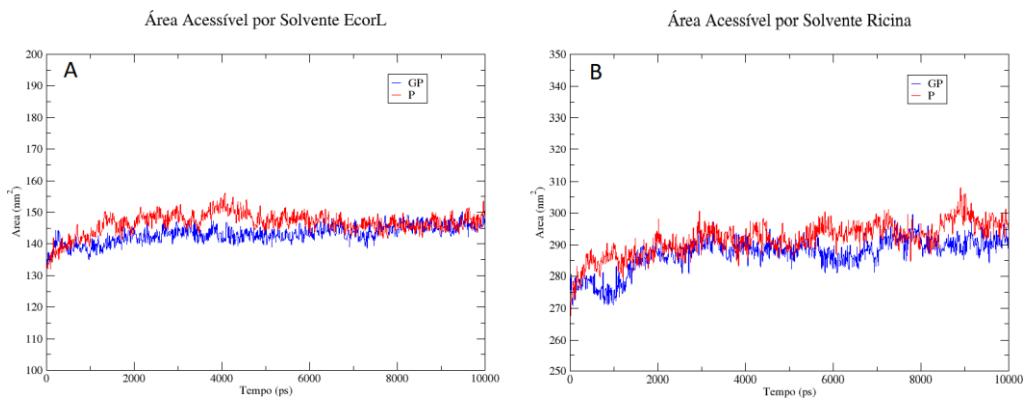
897

898 **Figura 10** – Presença de glicinas causando um aumento na flutuação dos resíduos: 61, 62 e
 899 64 da ML-I (A); 190, 191 e 196 da ML-I (B); 37 e 39 da Pha-L (C); 212 e 216 da Pha-L (D).

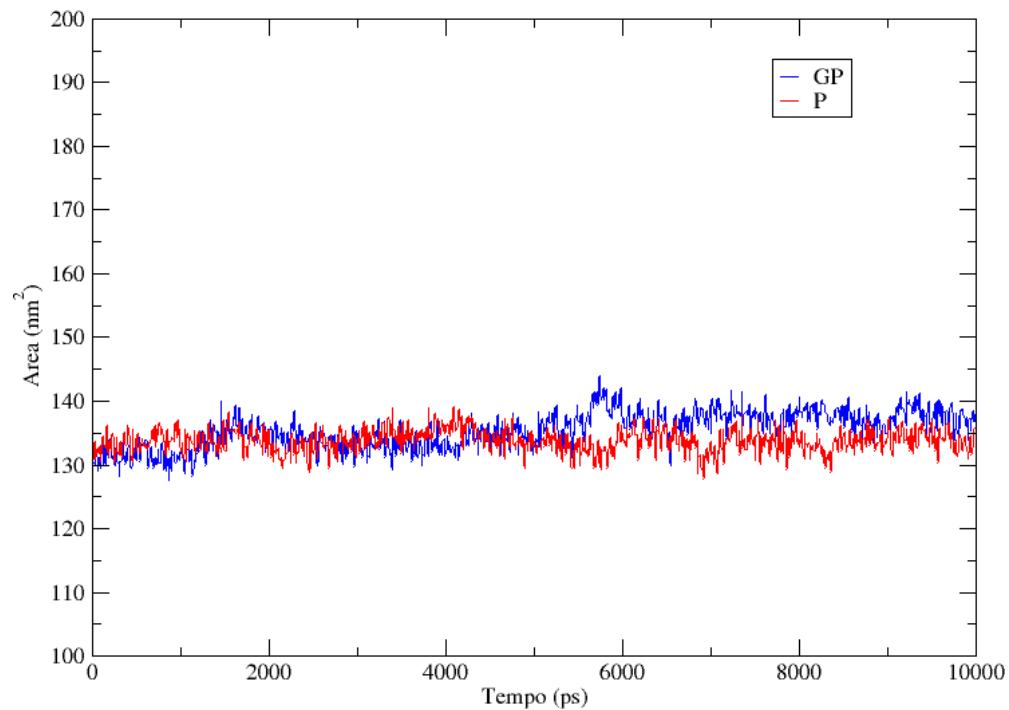
900

901

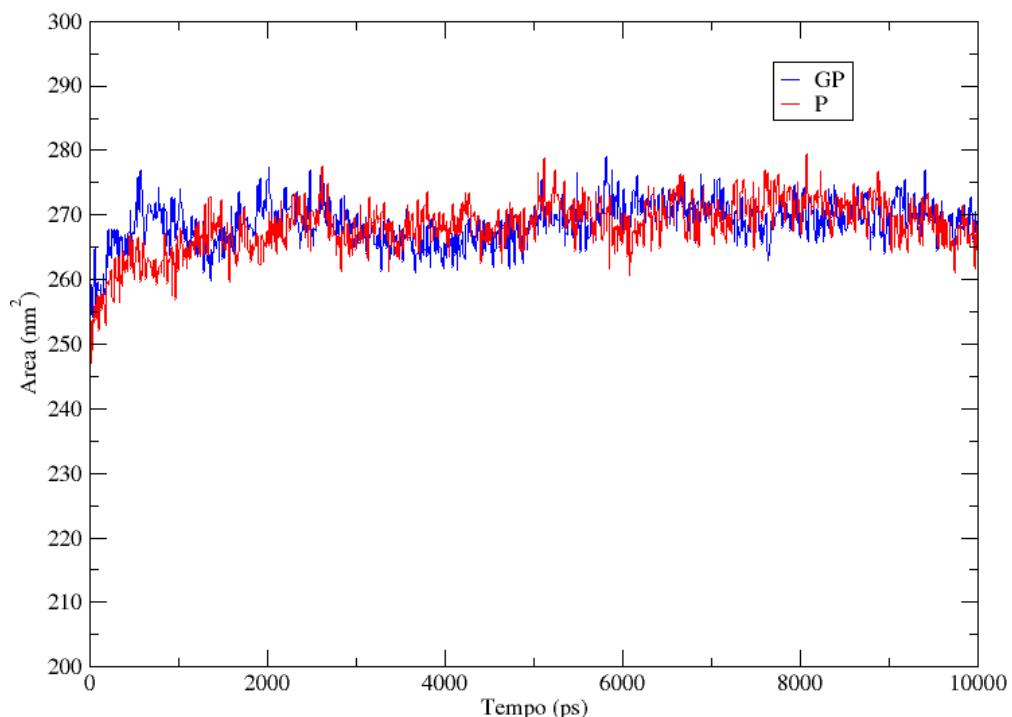
902 Na análise por área acessível de solvente, os resultados foram menos 903
 consistentes. As proteínas EcorL e Ricina apresentaram uma variação da área 904
 maior em suas versões não-glicosilada (Figura 11), enquanto a Pha-L obteve 905
 uma variação de área maior na versão glicosilada (Figura 12). Não houve 906
 diferença significativa na área da proteína ML-I (Figura 13).



Área Acessível por Solvente Pha-L



Área Acessível por Solvente ML-I



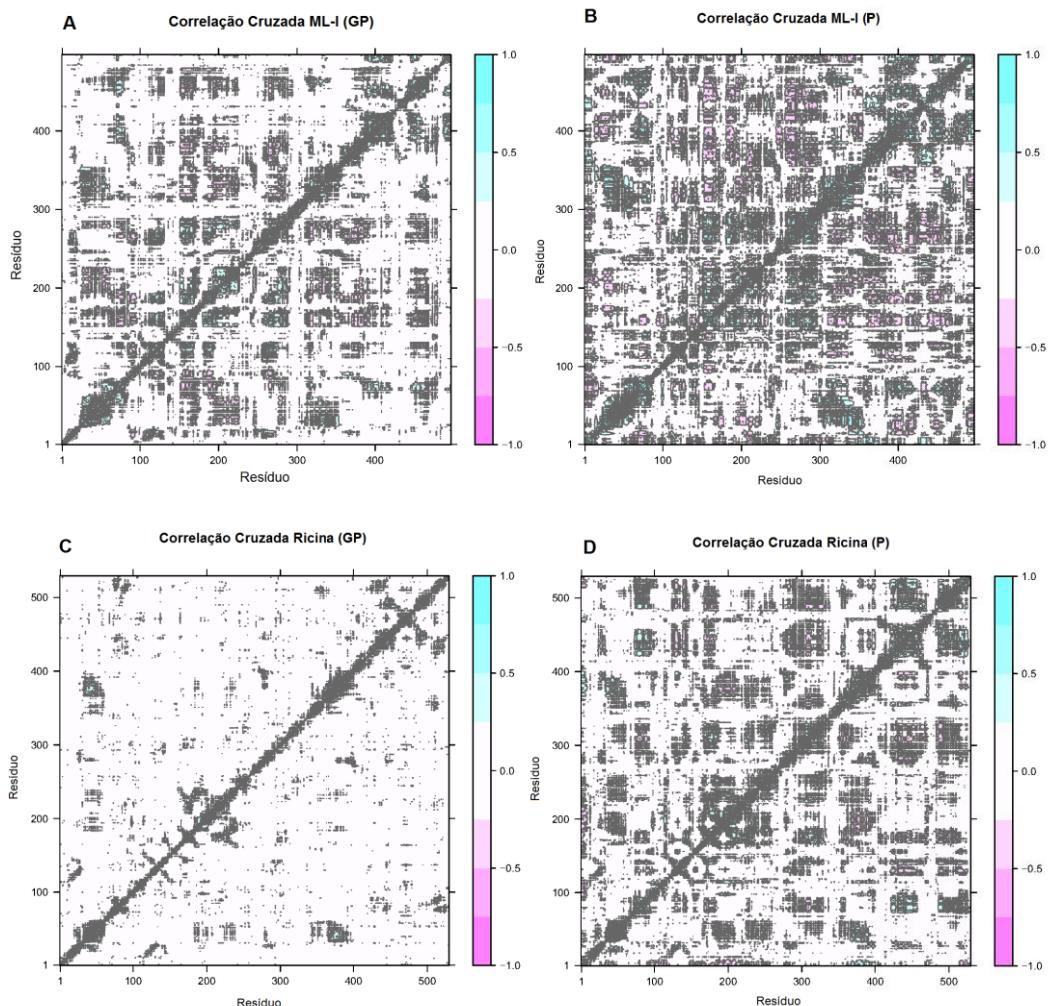
917

918 **Figura 13** – Área acessível por solvente da proteína ML-I em sua forma glicosilada (GP) e não-
 919 glicosilada (P). A média de área acessível da proteína GP foi de 268.75 ± 0.30 e da proteína P
 920 de 268.24 ± 0.25 ($P < 0.01$).

921

922

923 Do pacote Bio3D, a Análise de Correlação Cruzada (CCA, do inglês *Cross 924
 Correlation Analysis*) de resíduos também foi utilizada. Nessa análise, a 925
 interação dos resíduos é caracterizada por um espectro de cor indo do Azul (1) 926
 ao Vermelho (-1). Quanto mais positivo (azul), mais os resíduos se movem na 927
 mesma direção, enquanto quanto mais negativo (vermelho), mais os resíduos 928 se
 movem em direção oposta. Na Figura 14 podemos ver que há um menor 929 número
 de correlação e anti-correlação entre resíduos das proteínas ML-I e 930 Ricina em
 sua versão glicosilada, o que possivelmente significa uma menor 931 dinâmica
 interna da proteína.

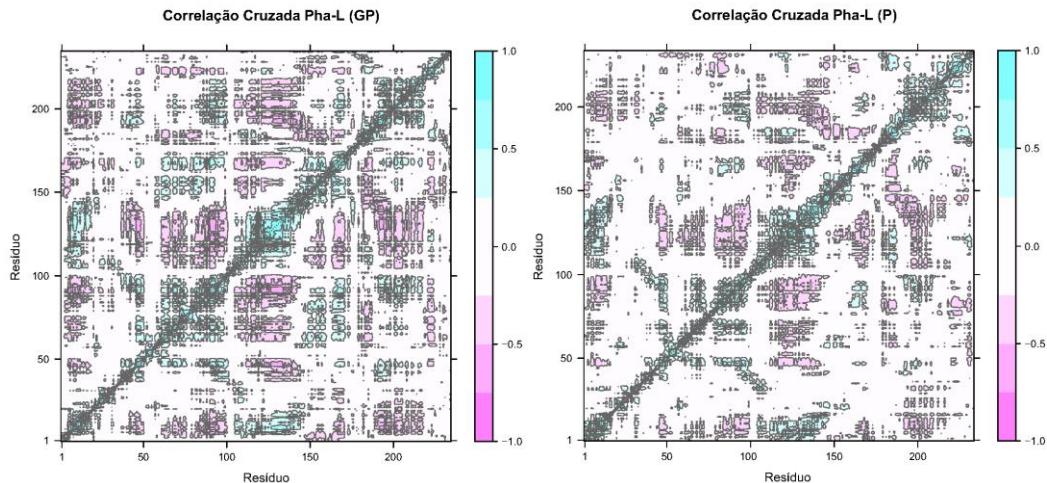


932

933 **Figura 14** - Análise de Correlação Cruzada das proteínas: ML-I Glicosilada (A) e não-
 934 glicosilada (B); Ricina Glicosilada (C) e não-glicosilada (D).

935

936 A proteína Pha-L , entretanto, teve um maior número de correlações e anti-
 937 correlações em sua forma não-glicosilada (Figura 15).

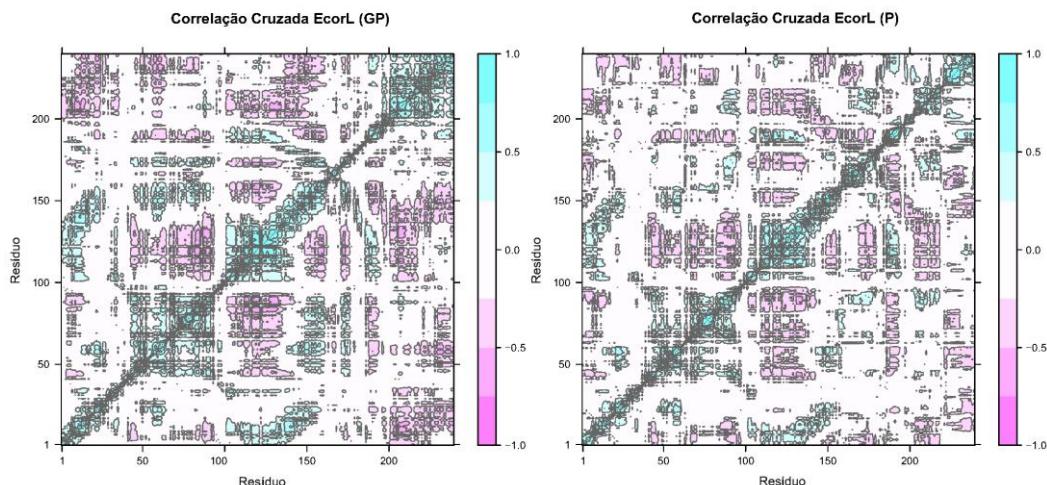


938

939 **Figura 15** - Análise de Correlação Cruzada da proteína Pha-L em sua forma glicosilada (GP) e
 940 não-glicosilada (P)

941

942 Já a proteína EcorL teve um número similar de correlações e anti-correlações
 943 em ambas as formas, o que pode demonstrar que as formas glicosiladas e não-
 944 glicosiladas possuem uma dinâmica interna similar (Figura 16).



945

946 **Figura 16** - Análise de Correlação Cruzada da proteína EcorL em sua forma glicosilada (GP) e
 947 não-glicosilada (P)

948

949

950

951

952 **5.2 Clusters**

953

954 Os *clusters* foram produzidos pelo método de GROMOS. Na Tabela 4 estão 955 disponíveis os valores de *cutoff* assim como o número de *clusters* gerados para 956 cada proteína. Já a Tabela 5 mostra o número de membros nos dois principais 957 Clusters de cada proteína.

958 Após gerados os *clusters*, a estrutura que representa o centróide de cada um 959 dos dois principais clusters de todas as proteínas foi selecionada para a 960 abordagem de *Docking*.

961

962 **Tabela 4** – Números de *clusters* formados por cada uma das proteínas em sua forma
963 glicosilada (GP) e não-glicosilada (P), assim como o seu respectivo valor de *cutoff*.

Proteína	PDB	Forma	RMSD <i>cutoff</i>	N. de Clusters
EcorL	1AXZ	GP	0.19 nm	7
		P	0.18 nm	4
ML-I	2MLL	GP	0.18 nm	3
		P	0.21 nm	3
Pha-L	1G8W	GP	0.15 nm	5
		P	0.16 nm	4
Ricina	2AAI	GP	0.17 nm	4
		P	0.19 nm	4

964

965

966

967

968

969

970

971

972

973 **Tabela 5** – Número de membros nos dois primeiros clusters das proteínas em sua forma
 974 glicosilada (GP) e não-glicosilada (P).

Proteína	PDB	Gorma	Cluster	População
EcorL	1AXZ	GP	1	871
			2	68
		P	1	764
			2	215
ML-I	2MLL	GP	1	785
			2	187
		P	1	819
			2	147
Pha-L	1G8W	GP	1	730
			2	235
		P	1	892
			2	79
Ricina	2AAI	GP	1	812
			2	154
		P	1	701
			2	214

975

976

977 **5.3 Docking Molecular**

978

979 Após produzidos os *clusters*, foi feita a abordagem de *docking* molecular. O 980
 981 sítio de ligação para cada proteína foi analizada pelo programa COACH. O 981
 982 centróide do primeiro cluster das proteínas em suas formas glicosiladas e não- 982
 983 glicosiladas foi docado, utilizando o sítio predito pelo coach, com seus ligantes 983
 984 nativos. No primeiro centróide, sendo o mais representativo, foram utilizados 984
 985 três programas de Docking para aferir o resultado da ligação. A Tabela 6 985 mostra
 986 os valores obtidos em cada programa.

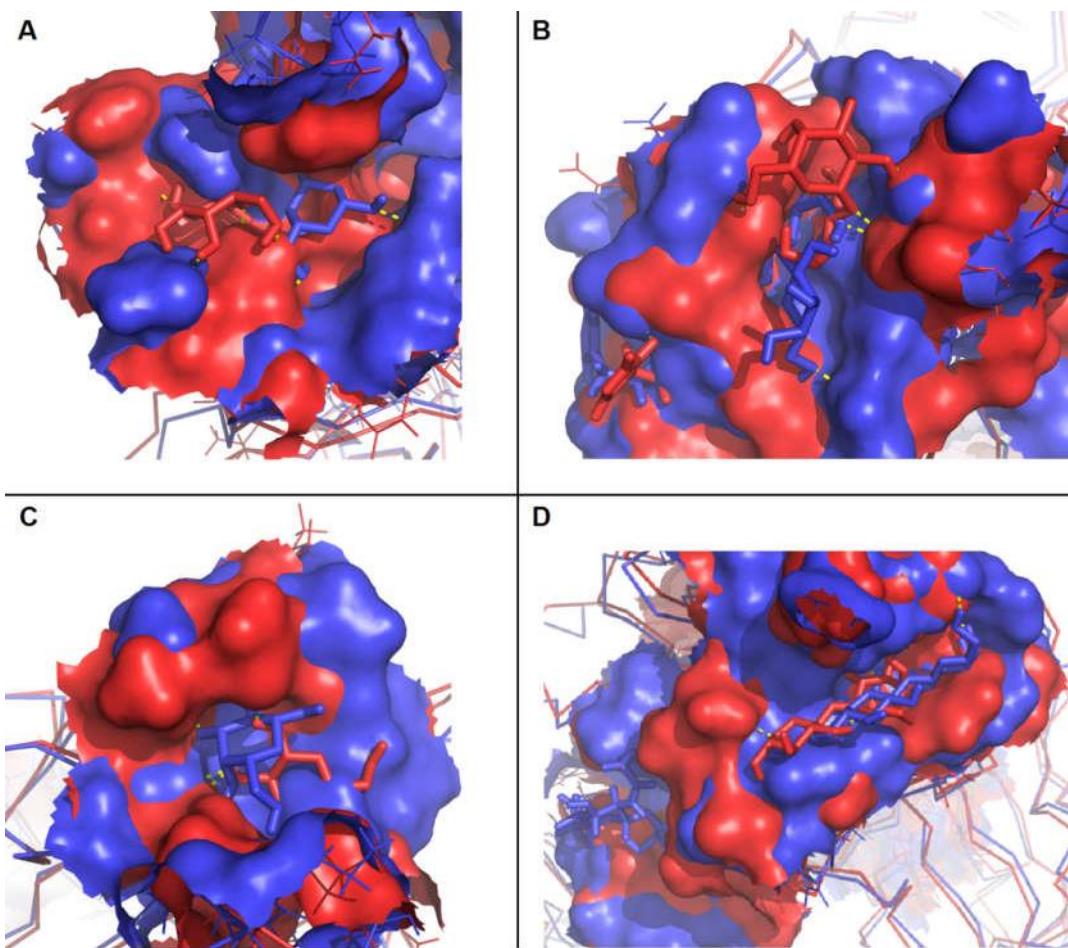
986

987
988 **Tabela 6** - Resultado dos programas de docking quanto a capacidade das proteínas em sua
989 forma glicosilada (GP) e não-glicosilada (P) de se ligarem a seu ligante. Em todos os casos,
quanto mais negativo o valor, maior o potencial da ligação.

Proteína	PDB	Ligante	Forma	Vina	CLC	SwissDock
EcorL	1AXZ	Galactose	GP	-4.2	-27.68	-6.97
			P	-3.8	-27.33	-6.09
ML-I	2MLL	Lactose	GP	-5.3	-35.29	-6.59
			P	-4.8	-29.59	-6.48
Pha-L	1G8W	Galactose	GP	-4.2	-26.35	-6.76
			P	-4.6	-36.99	-6.29
Ricina	2AAI	Lactose	GP	-5.7	-45.05	-6.37
			P	-5.2	-41.47	-6.16

990

991 As versões glicosiladas e não-glicosiladas das proteínas foram sobrepostas, 992
para mostrar a posição do ligante dentro do sítio de ligação da proteína (Figura 993
17). É possível observar que, apesar de eles não ocuparem exatamente o 994
mesmo espaço, devido ao movimento da proteína, ainda assim eles ocupam o 995
mesmo sítio.



996

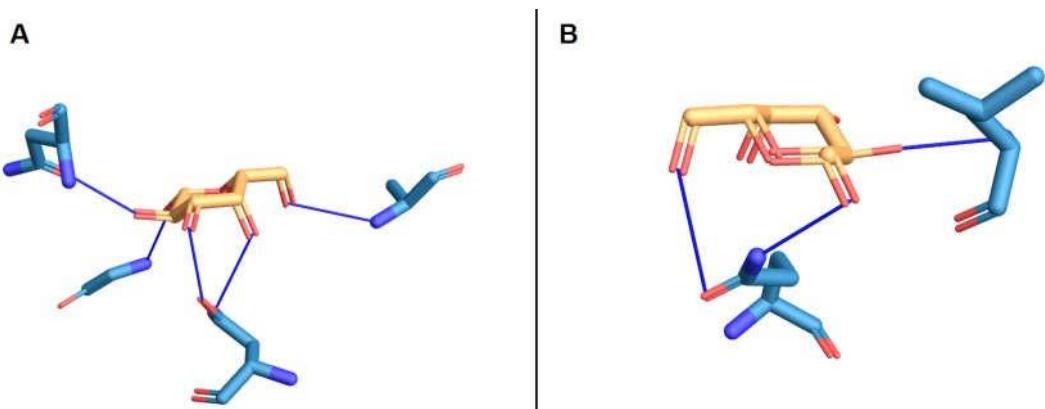
997 **Figura 17-** Sobreposição das proteínas EcorL (A), ML-I (B), Pha-L (C) e Ricina (D) juntamente
 998 com seus ligantes já docados, em suas versões glicosiladas (Azul) e não-glicosiladas
 999 (Vermelho)

1000

1001 Nas análises feitas usando o programa PLIP, foi possível observar o número de
 1002 pontes de hidrogênio e pontes salinas formadas pelas proteínas com seus
 1003 ligantes. Em ambas a EcorL (Figura 18) e ML-I (Figura 19) houve um aumento
 1004 no número de pontes de hidrogênio da proteína em sua forma glicosilada. Já
 1005 na proteína Pha-L (Figura 20), o número de pontes de hidrogênio foi igual para
 1006 as duas versões. Na Ricina (Figura 21), apesar do número de pontes de
 1007 hidrogênio ser igual, a versão não-glicosilada fez mais ligações salinas do que
 1008 a glicosilada. A Tabela 7 mostra o número de ligações assim como os resíduos
 1009 envolvidos na interação.

1010

1011

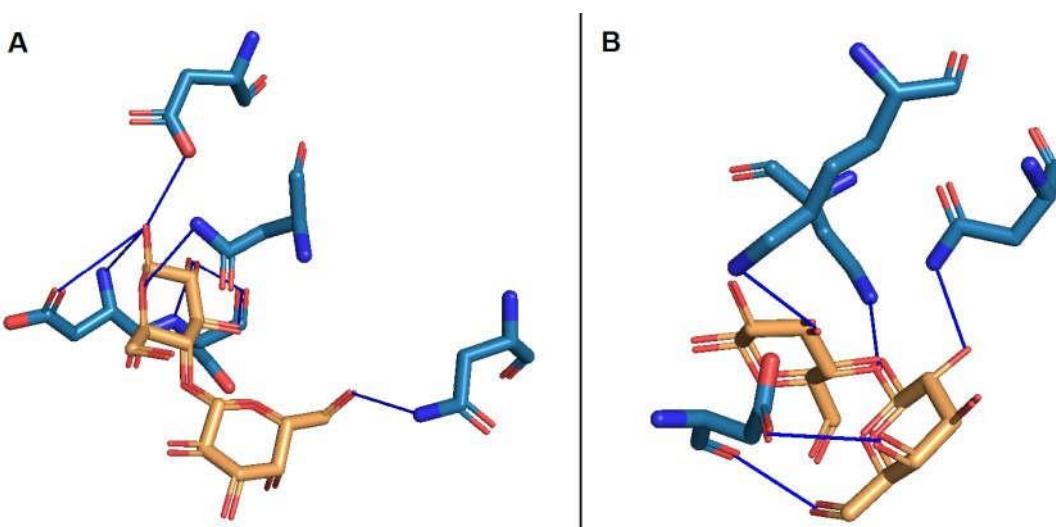


1012
1013 **Figura 18** - Ligações resultantes do docking da proteína EcorL em sua forma glicosilada (A) e
1014 não-glicosilada (B) com seu ligante (Galactose - Amarelo)

1015

1016

1017

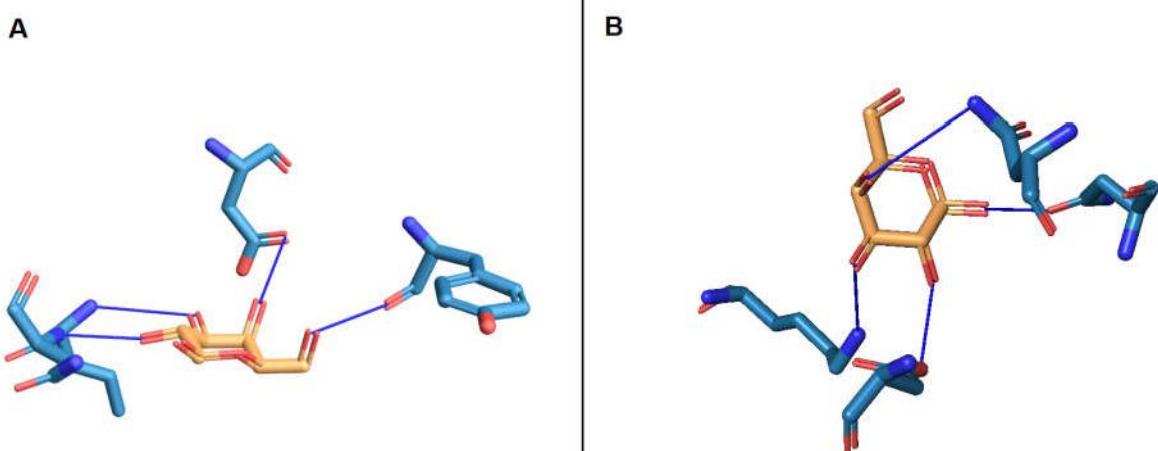


1018
1019 **Figura 19** - Ligações resultantes do docking da proteína ML-I em sua forma glicosilada (A) e
1020 não-glicosilada (B) com seu ligante (Lactose - Amarelo)

1021

1022

1023



1024

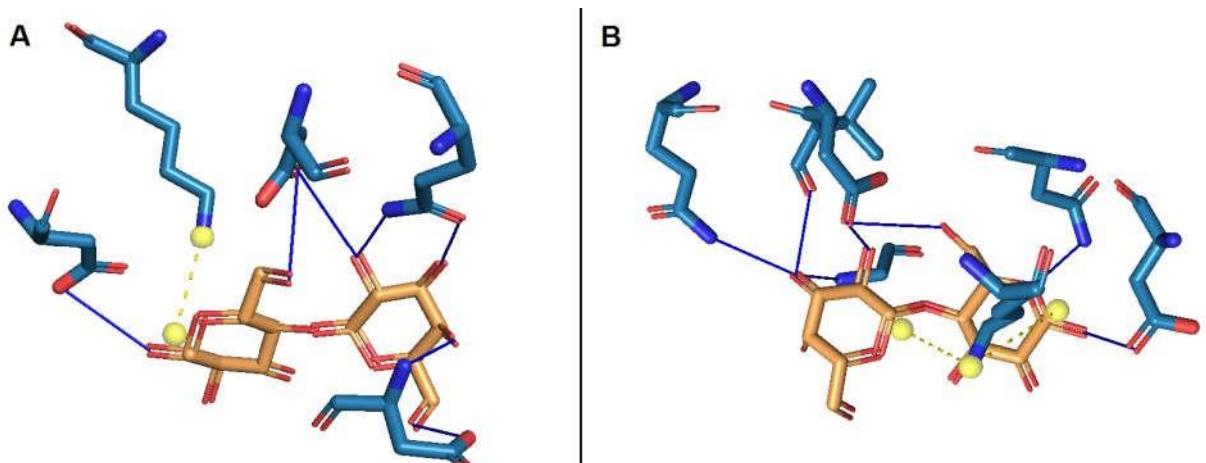
1025 **Figura 20** - Ligações resultantes do docking da proteína Pha-L em sua forma glicosilada (A) e
 1026 não-glicosilada (B) com seu ligante (Galactose - Amarelo)

1027

1028

1029

1030



1031

1032 **Figura 21** - Ligações resultantes do docking da proteína Ricina em sua forma glicosilada (A) e
 1033 não-glicosilada (B) com seu ligante (Lactose - Amarelo)

1034

1035

1036

1037

1038

1039

1040 **Tabela 7** - Número de pontes de hidrogênio e pontes salinas formadas pelo docking molecular
 1041 das proteínas em sua forma glicosilada (GP) e não-glicosilada (P), assim como os resíduos que
 1042 formam essas ligações.

Proteína	PDB	Forma	Pontes de		Resíduos
			Hidrogênio	Salinas	
EcorL	1AXZ	GP	5	-	89, 89, 107, 133, 218
		P	3	-	86, 219, 219 19, 23, 23, 24, 24, 24, 43, 57
ML-I	2MLL	GP	8	-	23, 23, 37, 39, 43
		P	5	-	86, 127, 213, 214
Pha-L	1G8W	GP	4	-	86, 129, 214, 217
		P	4	-	22, 22, 25, 25, 35, 35, 40, 44
Ricina	2AAI	GP	7	1	22, 22, 23, 26, 35, 40, 40, 44, 46
		P	7	2	

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058 **6 Discussão**

1059

1060 Tendo em vista o crescimento da produção de proteínas recombinantes, assim
1061 como as limitações em sistemas de expressão heterólogos quanto a
1062 glicosilação, é economicamente e cientificamente importante aferir os efeitos
1063 que essa modificação tem na atividade de proteínas, sendo previsto que mais
1064 de 50% das proteínas de Eucariotos sejam glicosiladas (ROTH; YEHEZKEL;
1065 KHALAILA, 2012). Este estudo é importante para botar em perspectiva a
1066 importância desse processo na produção de proteínas, e mostra que há um
1067 grande interesse em elucidar o funcionamento dessas glicoproteínas,
1068 especialmente a sua atividade biológica.

1069 Em todas as 4 proteínas estudadas, o RMSD foi menor nas proteínas
1070 glicosiladas do que não-glicosiladas (Tabela 2). Além disso, em todas, menos a
1071 EcorL, as proteínas glicosiladas demonstraram picos claros no histograma do
1072 RMSD, mostrando que ele se estabiliza em regiões específicas (Figura 2).
1073 Como o RMSD é obtido comparando as variações da estrutura com a estrutura
1074 nativa (cristal), essa análise nos mostra que as proteínas glicosiladas possuem
1075 uma maior estabilidade das estruturas, além de uma maior compactação.
1076 Quanto a EcorL (Figura 3), as diferenças nos valores da sua forma glicosilada
1077 muito provavelmente se deve ao pico irregular do RMSD por volta do 1ns
1078 (Figura 1a). Como esse ainda é um período em que a proteína está se
1079 estabilizando, essa diferença pode ser desconsiderada.

1080 Na análise de Raio de Giro (Figura 6), o resultado para todas as proteínas tanto
1081 na sua forma glicosilada quanto não-glicosilada foi igual. Nesse tipo de análise,
1082 o raio de giro tende a se manter estável se a proteína mantém o seu
1083 dobramento correto, e tende a variar quando ocorrem desdobramentos. Os
1084 resultados obtidos demonstram que não há diferenças no *folding* das proteínas
1085 glicosiladas e não-glicosiladas. Isso mostra que, apesar da glicosilação afetar a
1086 estrutura e a flutuação das proteínas, as estruturas mantêm o *folding* correto
1087 em ambas as formas.

1088 Nas análises por RMSF, é possível observar diferenças na dinâmica da
1089 proteína. A (Tabela 3) e a (Figura 7) mostram que a dinâmica dos resíduos é
1090 significativamente menor nas proteínas glicosiladas do que nas versões não-
1091 glicosiladas das proteínas EcorL, ML-I A, ML-I B e Ricina B ($P < 0.05$),
1092 mostrando uma maior estabilidade e corroborando o estudo realizado por
1093 (LEE; QI; IM, 2015). Entretanto não há significância estatística nas proteínas
1094 Pha-L ($P = 0.32$) e Ricina A ($P = 0.21$) (Figura 8). A (Figura 7) e (Figura 8)
1095 também mostram o local do sítio de glicosilação das proteínas e, juntamente
1096 com a (Figura 9), é possível perceber os maiores decréscimos nos picos de
1097 flutuação não necessariamente ocorrem nas regiões próximas ao glicano, o
1098 que mostra que o efeito da glicosilação não é localizado, mas se propaga por
1099 toda a cadeia da proteína. Esse fenômeno ocorre devido à modificações na
1100 molécula causarem perturbações que são propagadas para outras regiões da
1101 proteína, redistribuindo sua dinâmica para minimizar a perda de entropia
1102 causada pela modificação (GOHLKE; KUHN; CASE, 2004). É interessante
1103 notar que a Cadeia A da estrutura obtida por cristal da Ricina (que é um
1104 heterodímero) não possui glicano. Isso pode explicar os resultados obtidos,
1105 levando a teoria de que, apesar da redistribuição de perturbações causadas
1106 pela presença do glicano, possivelmente os efeitos da glicosilação de uma
1107 cadeia não significativamente afete a outra. Já a proteína Pha-L tem sua
1108 versão glicosilada normal e entretanto também apresentou resultados não
1109 significativos, não só no RMSF mas também em outras análises descritas
1110 abaixo.

1111 A (Figura 11) mostra que as proteínas EcorL e Ricina demonstraram ter uma
1112 área acessível por solvente menor em sua forma glicosilada. Uma área
1113 acessível menor também significa uma maior estabilidade da proteína, visto
1114 que um aumento na área acessível por solvente implica em um maior
1115 movimento de “abertura” nas regiões de cadeia lateral, causando mais
1116 instabilidade e aumento da dinâmica interna da proteína. A proteína ML-I
1117 (Figura 13) não demonstrou diferenças significativas entre as formas
1118 glicosiladas e não-glicosiladas, o que pode significar que não há diferenças
1119 significativas na estabilidade das regiões de cadeia lateral de ambas as formas
1120 da proteína. Já a Pha-L (Figura 12) demonstrou ter uma área acessível maior

1121 em sua forma glicosilada, o que, juntamente com a análise por RMSF,
1122 demonstra a instabilidade da forma glicosilada dessa proteína.

1123 Na correlação cruzada também podemos ver uma medida da dinâmica interna
1124 das proteínas, calculada através da interação entre os diferentes resíduos. Nas
1125 proteínas ML-I e Ricina (Figura 14) podemos ver uma interação muito maior
1126 entre os resíduos em suas formas não-glicosiladas do que da sua forma
1127 glicosilada, implicando em uma dinâmica interna menor nas glicoproteínas. É
1128 interessante notar que a Cadeia A da Ricina (Resíduos 1 a 267), apesar da
1129 análise por RMSF não ter demonstrado diferenças significativas entre as duas
1130 formas, mostrou possuir uma menor dinâmica interna na sua forma glicosilada.
1131 Já as proteínas Pha-L (Figura 15) e EcorL (Figura 16) demonstraram ter uma
1132 interação maior entre resíduos na sua forma glicosilada, o que implica em uma
1133 dinâmica interna maior na interação entre resíduos dessas glicoproteínas.

1134 Novamente a proteína Pha-L mostra uma estabilidade menor em sua forma
1135 glicosilada. Esse resultado somado aos anteriores mostra que essa proteína
1136 tem claramente uma estabilidade estrutural maior na forma não-glicosilada,
1137 indo contra à teoria geral de que proteínas glicosiladas vão possuir maior
1138 estabilidade estrutural. Há duas possibilidades de explicar esse resultado. Essa
1139 proteína pode ser de fato mais instável em sua forma glicosilada, pelo menos
1140 quanto a estabilidade geral da proteína, ou esses resultados podem ter sido
1141 causados por problemas na metodologia de dinâmica molecular. Um fator que
1142 corrobora a interpretação desses resultados terem sido causados por erros na
1143 metodologia é que a proteína foi simulada em sua forma monomérica,
1144 entretanto a atividade dela está ligado a formação de uma estrutura tetramérica
1145 (HAMELRYCK et al., 1996), o que pode ter comprometido a validade dos
1146 resultados.

1147 Quanto a clusterização da trajetória, é possível notar que, para todas as
1148 proteínas, mais de 90% da trajetória está contida nos dois primeiros clusters,
1149 sendo mais de 70% contido no primeiro (Tabela 5). Isso significa que, nas
1150 condições que foram gerados os clusters, um centróide obtido do primeiro
1151 clusters é representativo de cerca de 70% do movimento da estrutura no tempo
1152 simulado.

1153 As proteínas EcorL, ML-I e Ricina apresentaram um melhor resultado no
1154 docking da sua forma glicosilada em todos os 3 programas utilizados. Já a
1155 proteína Pha-L demonstrou um melhor potencial de ligação em sua forma não-
1156 glicosilada em 2 dos 3 programas utilizados (Tabela 6). Esses resultados estão
1157 alinhados com o que foi observado na maioria dos resultados da análise
1158 estrutural, com a proteína Pha-L, por exemplo, obtendo um melhor resultado
1159 em sua forma não-glicosilada do que na sua forma glicosilada. Apesar desse
1160 resultado ser contrário a hipótese de que proteínas glicosiladas tem um melhor
1161 potencial de ligação do que sua forma não-glicosilada, é importante observar
1162 que o resultado do *Docking* Molecular consegue mostrar que as diferenças da
1163 dinâmica da proteína em sua análise estrutural, principalmente RMSF, Área
1164 acessível por solvente e correlação cruzada, são refletidas no resultado do
1165 *docking*. Esse resultado é importante porque mostra que a diferenciação das
1166 proteínas por dinâmica molecular causa resultados observáveis por
1167 metodologias como *Docking*, o que torna viável o uso dessa combinação como
1168 uma *pipeline* de análise.

1169 É importante também observar que, em todas as proteínas, houve certas
1170 diferenças entre os locais de formação de ponte de hidrogênio entre as formas
1171 glicosiladas e não-glicosiladas (Figuras 18 a 21). Entretanto, como pode ser
1172 observado pela (Figura 17), os ligantes estão corretamente posicionados no
1173 sítio de ligação das proteínas. Uma forma de explicar essa diferença nas
1174 pontes de hidrogênio é devido as proteínas estarem em conformações
1175 diferentes por causa da sua movimentação. Alguns estudos mostram que
1176 modificações na posição da proteína causa uma diferença nos resíduos
1177 acessíveis a um ligante ou outra proteína (TOVCHIGRECHKO; WELLS;
1178 VAKSER, 2002), (JAIN, 2009).

1179 Das quatro proteínas estudadas, três delas – EcorL, ML-I e Ricina confirmaram
1180 a teoria de que proteínas glicosilada são mais estáveis do que sua forma não-
1181 glicosilada. A proteína Pha-L, entretanto, mostrou justamente o contrário. Mais
1182 estudos sobre essa proteína são necessários para definir se houve erro na
1183 metodologia, ou se é uma peculiaridade dessa proteína. A abordagem de
1184 *Docking* Molecular utilizando as estruturas obtidas através do método de
1185 *Clustering* seguiram o mesmo padrão da análise estrutural, obtendo valores

1186 favoráveis para as proteínas EcorL, ML-I e Ricina e valores não-favoráveis para
1187 a proteína Pha-L, demonstrando que modificações na estabilidade da proteína
1188 são refletidas na sua atividade, pelo menos no que tange sua ligação proteína-
1189 ligante.

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206 ‘

1207

1208

1209

1210 **7 Conclusão**

1211

1212 • É possível observar mudanças estruturais nas proteínas glicosiladas e
1213 não-glicosiladas após 10 nanosegundos de simulação. Apesar disso,
1214 ainda é ideal realizar os mesmos experimentos com um tempo maior de
1215 simulação, para confirmar se esses resultados vão se manter em um
1216 intervalo de tempo mais longo.

1217

1218 • Resultados mostram que Dinâmica Molecular seguida de *Docking* com
1219 receptor flexível é uma metodologia capaz de aferir as diferenças
1220 causadas pela presença ou ausência de um glicano na estrutura da
1221 proteína, validando a abordagem como uma *pipeline* de análise para
1222 estruturas glicosiladas.

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236 **8 Referências**

1237

1238 ABRAHAM, M. J. et al. GROMACS: High performance molecular simulations through
1239 multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 1
1240 set. 2015.

1241 ADCOCK, S. A.; MCCAMMON, J. A. Molecular dynamics: survey of methods for
1242 simulating the activity of proteins. **Chemical reviews**, v. 106, n. 5, p. 1589–615, maio
1243 2006.

1244 AKKOUH, O. et al. Lectins with Anti-HIV Activity: A Review. **Molecules**, v. 20, n. 1, p.
1245 648–668, 6 jan. 2015.

1246 ALLEN, W. J. et al. DOCK 6: Impact of new features and current docking performance.
1247 **Journal of Computational Chemistry**, v. 36, n. 15, p. 1132–1156, 5 jun. 2015.

1248 ALONSO, H.; BLIZNYUK, A. A.; GREADY, J. E. Combining docking and molecular
1249 dynamic simulations in drug design. **Medicinal Research Reviews**, v. 26, n. 5, p. 531–
1250 568, set. 2006.

1251 BERKHIN, P. A Survey of Clustering Data Mining Techniques. In: **Grouping
1252 Multidimensional Data**. Berlin/Heidelberg: Springer-Verlag, 2006. p. 25–71.

1253 BOTHWELL, J. H. F.; GRIFFIN, J. L. An introduction to biological nuclear magnetic
1254 resonance spectroscopy. **Biological Reviews**, v. 86, n. 2, p. 493–510, maio 2011.

1255 BROOIJMANS, N.; KUNTZ, I. D. Molecular Recognition and Docking Algorithms.
1256 **Annual Review of Biophysics and Biomolecular Structure**, v. 32, n. 1, p. 335–373,
1257 jun. 2003.

1258 BROOKS, S. A. Appropriate Glycosylation of Recombinant Proteins for Human Use:
1259 Implications of Choice of Expression System. **Molecular Biotechnology**, v. 28, n. 3, p.
1260 241–256, 2004.

1261 CHATZIELEFTHERIOU, S.; ADENDORFF, M. R.; LAGAROS, N. D. Generalized
1262 Potential Energy Finite Elements for Modeling Molecular Nanostructures. **Journal of
1263 Chemical Information and Modeling**, v. 56, n. 10, p. 1963–1978, 24 out. 2016.

1264 CHILDERS, M. C. et al. Insights from molecular dynamics simulations for
1265 computational protein design. **Mol. Syst. Des. Eng.**, v. 2, n. 1, p. 9–33, 2017.

1266 CORNELL, W. D. et al. A Second Generation Force Field for the Simulation of
1267 Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **1995**, 117,
1268 5179–5197. **Journal of the American Chemical Society**, v. 118, n. 9, p. 2309–2309,
1269 jan. 1996.

1270 COSCONATI, S. et al. Virtual Screening with AutoDock: Theory and Practice. **Expert
1271 opinion on drug discovery**, v. 5, n. 6, p. 597–607, 1 jun. 2010.

1272 COZZINI, P. et al. Target flexibility: an emerging consideration in drug discovery and
1273 design. **Journal of medicinal chemistry**, v. 51, n. 20, p. 6237–55, 23 out. 2008.

1274 DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \log(N)$ method
1275 for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p.
1276 10089–10092, 15 jun. 1993.

1277 DAURA, X. et al. Peptide Folding: When Simulation Meets Experiment. **Angewandte**
1278 **Chemie International Edition**, v. 38, n. 1–2, p. 236–240, 15 jan. 1999.

1279 DE PARIS, R. et al. Clustering molecular dynamics trajectories for optimizing docking
1280 experiments. **Computational Intelligence and Neuroscience**, v. 2015, p. 916240,
1281 2015.

1282 DELANO, W. L. **The PyMol Molecular Graphics System**. [s.l: s.n.]. v. 30

1283 DIAS, R. et al. Insights into Animal and Plant Lectins with Antimicrobial Activities.
1284 **Molecules**, v. 20, n. 1, p. 519–541, 5 jan. 2015.

1285 ELGAVISH, S.; SHAANAN, B. Structures of the Erythrina corallodendron lectin and of
1286 its complexes with mono- and disaccharides. **Journal of molecular biology**, v. 277, n.
1287 4, p. 917–32, 10 abr. 1998.

1288 ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial
1289 databases with noise. p. 226--231, 1996.

1290 FORLI, S. Charting a Path to Success in Virtual Screening. **Molecules (Basel, Switzerland)**, v. 20, n. 10, p. 18732–58, 15 out. 2015.

1292 FRACCALVIERI, D. et al. Conformational and functional analysis of molecular
1293 dynamics trajectories by self-organising maps. **BMC bioinformatics**, v. 12, p. 158, 14
1294 maio 2011.

1295 FRIESNER, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring.
1296 1. Method and assessment of docking accuracy. **Journal of medicinal chemistry**, v.
1297 47, n. 7, p. 1739–49, 25 mar. 2004.

1298 GALLO, M. T. et al. Novel procedure for thermal equilibration in molecular dynamics
1299 simulation. **Molecular simulation**, v. 35, n. 5, p. 349–357, 1 abr. 2009.

1300 GARBUZYNSKIY, S. O. et al. Comparison of X-ray and NMR structures: Is there a
1301 systematic difference in residue contacts between X-ray- and NMR-resolved protein
1302 structures? **Proteins: Structure, Function, and Bioinformatics**, v. 60, n. 1, p. 139–
1303 147, 26 abr. 2005.

1304 GAVRILOV, B. K. et al. Effects of glycosylation on antigenicity and immunogenicity of
1305 classical swine fever virus envelope proteins. **Virology**, v. 420, n. 2, p. 135–145, 25
1306 nov. 2011.

1307 GIOIA, D. et al. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery.
1308 **Molecules**, v. 22, n. 12, p. 2029, 22 nov. 2017.

1309 GOETTIG, P. Effects of Glycosylation on the Enzymatic Activity and Mechanisms of
1310 Proteases. **International journal of molecular sciences**, v. 17, n. 12, 25 nov. 2016.

1311 GOHLKE, H.; KUHN, L. A.; CASE, D. A. Change in protein flexibility upon complex
1312 formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework
1313 approach. **Proteins: Structure, Function, and Bioinformatics**, v. 56, n. 2, p. 322–
1314 337, 14 maio 2004.

1315 GOOCHEE, C. F. et al. The Oligosaccharides of Glycoproteins: Bioprocess Factors
1316 Affecting Oligosaccharide Structure and their Effect on Glycoprotein Properties.
1317 **Bio/Technology**, v. 9, n. 12, p. 1347–1355, 1 dez. 1991.

1318 GOVINDARAJAN, S.; GOLDSTEIN, R. A. On the thermodynamic hypothesis of protein

1319 folding. **Proceedings of the National Academy of Sciences of the United States of**
 1320 **America**, v. 95, n. 10, p. 5545–9, 12 maio 1998.

1321 GRACE, M. E.; GRABOWSKI, G. A. Human acid beta-glucosidase: glycosylation is
 1322 required for catalytic activity. **Biochemical and biophysical research**
 1323 **communications**, v. 168, n. 2, p. 771–7, 30 abr. 1990.

1324 GRANT, B. J. et al. Bio3d: an R package for the comparative analysis of protein
 1325 structures. **Bioinformatics**, v. 22, n. 21, p. 2695–2696, 1 nov. 2006.

1326 GREENE, E. R. et al. Chapter Three – Glycosylation of Cellulases: Engineering Better
 1327 Enzymes for Biofuels. In: **Advances in Carbohydrate Chemistry and Biochemistry**.
 1328 [s.l: s.n.]. v. 72p. 63–112.

1329 GRINTER, S. Z.; ZOU, X. Challenges, applications, and recent advances of protein-
 1330 ligand docking in structure-based drug design. **Molecules (Basel, Switzerland)**, v. 19,
 1331 n. 7, p. 10150–76, jan. 2014.

1332 GROS DIDIER, A.; ZOETE, V.; MICHELIN, O. SwissDock, a protein-small molecule
 1333 docking web service based on EADock DSS. **Nucleic Acids Research**, v. 39, n. suppl,
 1334 p. W270–W277, 1 jul. 2011.

1335 HAMELRYCK, T. W. et al. The crystallographic structure of phytohemagglutinin-L. **The**
 1336 **Journal of biological chemistry**, v. 271, n. 34, p. 20479–85, 23 ago. 1996.

1337 HAMID, R. et al. Lectins: Proteins with diverse applications. **Journal of Applied**
 1338 **Pharmaceutical Science**, v. 3, n. 4SUPPL.1, 2013.

1339 HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. **Journal of**
 1340 **Computational Chemistry**, v. 18, n. 12, p. 1463–1472, 1 set. 1997.

1341 HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. **Physical**
 1342 **Review A**, v. 31, n. 3, p. 1695–1697, 1 mar. 1985.

1343 HOSPITAL, A. et al. Molecular dynamics simulations: advances and applications.
 1344 **Advances and applications in bioinformatics and chemistry: AABC**, v. 8, p. 37–
 1345 47, 2015.

1346 HUANG, J.; MACKERELL, A. D. CHARMM36 all-atom additive protein force field:
 1347 Validation based on comparison to NMR data. **Journal of Computational Chemistry**,
 1348 v. 34, n. 25, p. 2135–2145, 30 set. 2013.

1349 HUANG, S.-Y.; GRINTER, S. Z.; ZOU, X. Scoring functions and their evaluation
 1350 methods for protein–ligand docking: recent advances and future directions. **Physical**
 1351 **Chemistry Chemical Physics**, v. 12, n. 40, p. 12899, 7 out. 2010.

1352 ILARI, A.; SAVINO, C. Protein Structure Determination by X-Ray Crystallography. In:
 1353 **Methods in molecular biology (Clifton, N.J.)**. [s.l: s.n.]. v. 452p. 63–87.

1354 JACOB, J.; DUCLOHIER, H.; CAFISO, D. S. The Role of Proline and Glycine in
 1355 Determining the Backbone Flexibility of a Channel-Forming Peptide. **Biophysical**
 1356 **Journal**, v. 76, n. 3, p. 1367–1376, mar. 1999.

1357 JAIN, A. N. Effects of protein conformation in docking: improved pose prediction
 1358 through protein pocket adaptation. **Journal of computer-aided molecular design**, v.
 1359 23, n. 6, p. 355–74, jun. 2009.

1360 JAKOPITSCH, C. et al. Role of the Main Access Channel of Catalase-Peroxidase in

1361 Catalysis. **Journal of Biological Chemistry**, v. 280, n. 51, p. 42411–42422, 23 dez.
1362 2005.

1363 JO, S. et al. Glycan Reader: automated sugar identification and simulation preparation
1364 for carbohydrates and glycoproteins. **Journal of computational chemistry**, v. 32, n.
1365 14, p. 3135–41, 15 nov. 2011.

1366 JONES, G. et al. Development and validation of a genetic algorithm for flexible docking
1367 1 Edited by F. E. Cohen. **Journal of Molecular Biology**, v. 267, n. 3, p. 727–748, 4
1368 abr. 1997.

1369 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and Testing
1370 of the OPLS All-Atom Force Field on Conformational Energetics and Properties of
1371 Organic Liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225–
1372 11236, jan. 1996.

1373 KANDT, C.; ASH, W. L.; PETER TIELEMAN, D. Setting up and running molecular
1374 dynamics simulations of membrane proteins. **Methods**, v. 41, n. 4, p. 475–488, 1 abr.
1375 2007.

1376 KAUSHAL, G. P.; SZUMILO, T.; ELBEIN, A. D. Structure and Biosynthesis of Plant N-
1377 Linked Glycoproteins. In: **The Biochemistry of Plants**. [s.l.] Elsevier, 1988. p. 421–
1378 463.

1379 KELLER, B.; DAURA, X.; VAN GUNSTEREN, W. F. Comparing geometric and kinetic
1380 cluster algorithms for molecular simulation data. **The Journal of Chemical Physics**, v.
1381 132, n. 7, p. 74110, 21 fev. 2010a.

1382 KELLER, B.; DAURA, X.; VAN GUNSTEREN, W. F. Comparing geometric and kinetic
1383 cluster algorithms for molecular simulation data. **The Journal of Chemical Physics**, v.
1384 132, n. 7, p. 74110, 21 fev. 2010b.

1385 KENDREW, J. C. et al. A Three-Dimensional Model of the Myoglobin Molecule
1386 Obtained by X-Ray Analysis. **Nature**, v. 181, n. 4610, p. 662–666, 8 mar. 1958.

1387 KHOURY, G. A.; BALIBAN, R. C.; FLOUDAS, C. A. Proteome-wide post-translational
1388 modification statistics: frequency analysis and curation of the swiss-prot database.
1389 **Scientific Reports**, v. 1, n. 1, p. 90, 13 dez. 2011.

1390 KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery:
1391 methods and applications. **Nature reviews. Drug discovery**, v. 3, n. 11, p. 935–49,
1392 nov. 2004.

1393 KLAFKE, G. B. et al. Lectin I from Bauhinia variegata (BVL-I) expressed by Pichia
1394 pastoris inhibits initial adhesion of oral bacteria in vitro. **International Journal of**
1395 **Biological Macromolecules**, v. 93, p. 913–918, 1 dez. 2016.

1396 KOSHLAND, D. E. CORRELATION OF STRUCTURE AND FUNCTION IN ENZYME
1397 ACTION. **Science (New York, N.Y.)**, v. 142, n. 3599, p. 1533–41, 20 dez. 1963.

1398 KRAUSPENHAAR, R. et al. Crystal Structure of Mistletoe Lectin I from Viscum album.
1399 **Biochemical and Biophysical Research Communications**, v. 257, n. 2, p. 418–424,
1400 abr. 1999.

1401 KRISHNAN, V. et al. Macromolecular Structure Determination: Comparison of X-ray
1402 Crystallography and NMR Spectroscopy. In: **eLS**. Chichester, UK: John Wiley & Sons,
1403 Ltd, 2012.

1404 KUMAR, S. et al. Folding and binding cascades: dynamic landscapes and population
 1405 shifts. **Protein science: a publication of the Protein Society**, v. 9, n. 1, p. 10–9, jan.
 1406 2000.

1407 KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions.
 1408 **Journal of molecular biology**, v. 161, n. 2, p. 269–88, 25 out. 1982.

1409 LANNOO, N.; VAN DAMME, E. J. M. Lectin domains at the frontiers of plant defense.
 1410 **Frontiers in plant science**, v. 5, p. 397, 2014.

1411 LEE, H. S.; QI, Y.; IM, W. Effects of N-glycosylation on protein conformation and
 1412 dynamics: Protein Data Bank analysis and molecular dynamics simulation study.
 1413 **Scientific reports**, v. 5, p. 8926, 9 mar. 2015.

1414 LIU, D. T.-Y. Glycoprotein pharmaceuticals: scientific and regulatory considerations,
 1415 and the US Orphan Drug Act. **Trends in Biotechnology**, v. 10, p. 114–120, 1992.

1416 MACKERELL, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and
 1417 Dynamics Studies of Proteins †. **The Journal of Physical Chemistry B**, v. 102, n. 18,
 1418 p. 3586–3616, 30 abr. 1998.

1419 MARGRAF-SCHÖNFELD, S.; BÖHM, C.; WATZL, C. Glycosylation affects ligand
 1420 binding and function of the activating natural killer cell receptor 2B4 (CD244) protein.
 1421 **The Journal of biological chemistry**, v. 286, n. 27, p. 24142–9, 8 jul. 2011.

1422 MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. **Nature**,
 1423 v. 267, n. 5612, p. 585–590, 16 jun. 1977.

1424 MICHELETTI, C. Comparing proteins by their internal dynamics: Exploring structure–
 1425 function relationships beyond static structural alignments. **Physics of Life Reviews**, v.
 1426 10, n. 1, p. 1–26, mar. 2013.

1427 MILLS, C. L.; BEUNING, P. J.; ONDRECHEN, M. J. Biochemical functional predictions
 1428 for protein structures of unknown or uncertain function. **Computational and structural**
 1429 **biotechnology journal**, v. 13, p. 182–91, 2015.

1430 MIMURA, Y. et al. The influence of glycosylation on the thermal stability and effector
 1431 function expression of human IgG1-Fc: properties of a series of truncated glycoforms.
 1432 **Molecular Immunology**, v. 37, n. 12–13, p. 697–706, 1 set. 2000.

1433 MISHRA, V. et al. Purification and characterization of four isoforms of Himalayan
 1434 mistletoe ribosome-inactivating protein from Viscum album having unique sugar affinity.
 1435 **Archives of Biochemistry and Biophysics**, v. 423, n. 2, p. 288–301, 15 mar. 2004.

1436 MITRA, N. et al. N-linked oligosaccharides as outfitters for glycoprotein folding, form
 1437 and function. **Trends in Biochemical Sciences**, v. 31, n. 3, p. 156–163, mar. 2006.

1438 MOENS, S.; VANDERLEYDEN, J. Glycoproteins in prokaryotes. **Archives of**
 1439 **microbiology**, v. 168, n. 3, p. 169–75, set. 1997.

1440 NETO, L. G. DO N. et al. Effect of the Lectin of Bauhinia variegata and Its
 1441 Recombinant Isoform on Surgically Induced Skin Wounds in a Murine Model.
 1442 **Molecules**, v. 16, n. 12, p. 9298–9315, 7 nov. 2011.

1443 NOTHAFT, H.; SZYMANSKI, C. M. Protein glycosylation in bacteria: sweeter than
 1444 ever. **Nature Reviews Microbiology**, v. 8, n. 11, p. 765–778, nov. 2010.

1445 OLIVEIRA, C.; TEIXEIRA, J. A.; DOMINGUES, L. Recombinant lectins: an array of

1446 tailor-made glycan-interaction biosynthetic tools.
 1447 <http://dx.doi.org/10.3109/07388551.2012.670614>, 2013.

1448 OLIVEIRA, C.; TEIXEIRA, J. A.; DOMINGUES, L. Recombinant production of plant
 1449 lectins in microbial systems for biomedical application - the frutalin case study.
 1450 **Frontiers in plant science**, v. 5, p. 390, 2014.

1451 OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of
 1452 hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6.
 1453 **Journal of Computational Chemistry**, v. 25, n. 13, p. 1656–1676, out. 2004.

1454 PAGADALA, N. S.; SYED, K.; TUSZYNSKI, J. Software for molecular docking: a
 1455 review. **Biophysical reviews**, v. 9, n. 2, p. 91–102, abr. 2017.

1456 PAQUET, E.; VIKTOR, H. L. Molecular dynamics, monte carlo simulations, and
 1457 langevin dynamics: a computational review. **BioMed research international**, v. 2015,
 1458 p. 183918, 2015.

1459 PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new
 1460 molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182–7190,
 1461 14 dez. 1981.

1462 PATODIA, S.; BAGARIA, A.; CHOPRA, D. Molecular Dynamics Simulation of Proteins:
 1463 A Brief Overview. **Journal of Physical Chemistry & Biophysics**, v. 4, n. 6, 2014.

1464 PAULSON, J. C.; COLLEY, K. J. Glycosyltransferases. Structure, localization, and
 1465 control of cell type-specific glycosylation. **The Journal of biological chemistry**, v.
 1466 264, n. 30, p. 17615–8, 25 out. 1989.

1467 PERARO, M. D. et al. Modeling the charge distribution at metal sites in proteins for
 1468 molecular dynamics simulations. **Journal of Structural Biology**, v. 157, n. 3, p. 444–
 1469 453, 1 mar. 2007.

1470 RICHARDSON, P. T. et al. The effects of N-glycosylation on the lectin activity of
 1471 recombinant ricin B chain. **Carbohydrate Research**, v. 213, p. 19–25, jun. 1991.

1472 ROSENFELD, R.; VAJDA, S.; DELISI, C. FLEXIBLE DOCKING AND DESIGN. **Annu.**
 1473 **Rev. Biophys. Biomol. Struct.**, v. 24, p. 677–700, 1995.

1474 ROTH, Z.; YEHEZKEL, G.; KHALAILA, I. Identification and Quantification of Protein
 1475 Glycosylation. **International Journal of Carbohydrate Chemistry**, v. 2012, p. 1–10, 5
 1476 abr. 2012.

1477 ROUDI, Y.; AURELL, E.; HERTZ, J. A. Statistical physics of pairwise probability
 1478 models. **Frontiers in computational neuroscience**, v. 3, p. 22, 2009.

1479 RUTENBER, E. et al. Crystallographic refinement of ricin to 2.5 Å. **Proteins: Structure, Function, and Genetics**, v. 10, n. 3, p. 240–250, mar. 1991.

1481 SALENTIN, S. et al. PLIP: fully automated protein–ligand interaction profiler. **Nucleic**
 1482 **Acids Research**, v. 43, n. W1, p. W443–W447, 1 jul. 2015.

1483 SHAO, J. et al. Clustering Molecular Dynamics Trajectories: 1. Characterizing the
 1484 Performance of Different Clustering Algorithms. **Journal of Chemical Theory and**
 1485 **Computation**, v. 3, n. 6, p. 2312–2334, nov. 2007.

1486 SHARON, N.; LIS, H. Lectins as cell recognition molecules. **Science**, n. 246, p. 227–
 1487 234, 1989.

1488 SHENKIN, P. S.; MCDONALD, D. Q. Cluster analysis of molecular conformations.
1489 **Journal of Computational Chemistry**, v. 15, n. 8, p. 899–916, 1 ago. 1994.

1490 SHENTAL-BECHOR, D.; LEVY, Y. Effect of glycosylation on protein folding: a close
1491 look at thermodynamic stabilization. **Proceedings of the National Academy of**
1492 **Sciences of the United States of America**, v. 105, n. 24, p. 8256–61, 17 jun. 2008.

1493 SHI, X.; JARVIS, D. L. Protein N-glycosylation in the baculovirus-insect cell system.
1494 **Current drug targets**, v. 8, n. 10, p. 1116–25, out. 2007.

1495 SHI, Y. A glimpse of structural biology through X-ray crystallography. **Cell**, v. 159, n. 5,
1496 p. 995–1014, 20 nov. 2014.

1497 SINGH, S.; DEVI, S.; NG, T. Banana Lectin: A Brief Review. **Molecules**, v. 19, n. 11, p.
1498 18817–18827, 17 nov. 2014.

1499 SOLÁ, R. J.; GRIEBENOW, K. Effects of glycosylation on the stability of protein
1500 pharmaceuticals. **Journal of pharmaceutical sciences**, v. 98, n. 4, p. 1223–45, abr.
1501 2009.

1502 SOLÁ, R. J.; RODRÍGUEZ-MARTÍNEZ, J. A.; GRIEBENOW, K. Modulation of protein
1503 biophysical properties by chemical glycosylation: biochemical insights and biomedical
1504 implications. **Cellular and Molecular Life Sciences**, v. 64, n. 16, p. 2133–2152, 11
1505 ago. 2007.

1506 STEINBACH, P. J.; BROOKS, B. R. New spherical-cutoff methods for long-range
1507 forces in macromolecular simulation. **Journal of Computational Chemistry**, v. 15, n.
1508 7, p. 667–683, 1 jul. 1994.

1509 TORDA, A. E.; VAN GUNSTEREN, W. F. Algorithms for clustering molecular dynamics
1510 configurations. **Journal of Computational Chemistry**, v. 15, n. 12, p. 1331–1340, 1
1511 dez. 1994.

1512 TOVCHIGRECHKO, A.; WELLS, C. A.; VAKSER, I. A. Docking of protein models.
1513 **Protein science: a publication of the Protein Society**, v. 11, n. 8, p. 1888–96, ago.
1514 2002.

1515 TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of
1516 docking with a new scoring function, efficient optimization, and multithreading. **Journal**
1517 **of computational chemistry**, v. 31, n. 2, p. 455–61, 30 jan. 2010.

1518 TROYER, J. M.; COHEN, F. E. Protein conformational landscapes: energy
1519 minimization and clustering of a long molecular dynamics trajectory. **Proteins**, v. 23, n.
1520 1, p. 97–110, set. 1995.

1521 TSAI, C. J. et al. Folding funnels, binding funnels, and protein function. **Protein**
1522 **science: a publication of the Protein Society**, v. 8, n. 6, p. 1181–90, jun. 1999.

1523 VAHL QUEVEDO, C. et al. A strategic solution to optimize molecular docking
1524 simulations using Fully-Flexible Receptor models. **Expert Systems with Applications**,
1525 v. 41, n. 16, p. 7608–7620, 15 nov. 2014.

1526 VAKSER, I. A. Protein-protein docking: from interaction to interactome. **Biophysical**
1527 **journal**, v. 107, n. 8, p. 1785–93, 21 out. 2014.

1528 VAN PARIJS, J. et al. Hevein: an antifungal protein from rubber-tree (*Hevea*
1529 *brasiliensis*) latex. **Planta**, v. 183, n. 2, p. 258–264, jan. 1991.

1530 WALSH, C. T.; GARNEAU-TSODIKOVA, S.; GATTO, G. J. Protein Posttranslational
1531 Modifications: The Chemistry of Proteome Diversifications. **Angewandte Chemie**
1532 **International Edition**, v. 44, n. 45, p. 7342–7372, 18 nov. 2005.

1533 WEIS, W. I.; DRICKAMER, K. Structural Basis of Lectin-Carbohydrate Recognition.
1534 **Annual Review of Biochemistry**, v. 65, n. 1, p. 441–473, jun. 1996.

1535 WOLF, A.; KIRSCHNER, K. N. Principal component and clustering analysis on
1536 molecular dynamics data of the ribosomal L11-23S subdomain. **Journal of molecular**
1537 **modeling**, v. 19, n. 2, p. 539–49, fev. 2013.

1538 XU, Y.-H. et al. Two novel C-type lectins with a low-density lipoprotein receptor class A
1539 domain have antiviral function in the shrimp *Marsupenaeus japonicus*. **Developmental**
1540 **& Comparative Immunology**, v. 42, n. 2, p. 323–332, 2014.

1541 YANG, J.; ROY, A.; ZHANG, Y. Protein-ligand binding site recognition using
1542 complementary binding-specific substructure comparison and sequence profile
1543 alignment. **Bioinformatics (Oxford, England)**, v. 29, n. 20, p. 2588–95, 15 out. 2013.

1544 YAU, T. et al. Lectins with Potential for Anti-Cancer Therapy. **Molecules**, v. 20, n. 3, p.
1545 3791–3810, 26 fev. 2015.

1546 ZHANG, X.; WANG, Y. Glycosylation Quality Control by the Golgi Structure. **Journal of**
1547 **molecular biology**, v. 428, n. 16, p. 3183–3193, 14 ago. 2016.

1548