

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**Estudos computacionais do modelo de mucina 2 e sua
interação com quitosanas tioladas: uma nova estratégia
para o estudo da mucoadesão e retenção de fármacos**

*Computational studies of mucin 2 model and its interaction with thiolated
chitosans: a new insight for mucus adhesion and drug retention*

Cleverton Oliveira Cavalcanti da Silveira

Dissertação de Mestrado em Química

**Vitória
2018**

Cleverton Oliveira Cavalcanti da Silveira

**Estudos computacionais do modelo de mucina 2 e sua interação
com quitosanas tioladas: uma nova estratégia para o estudo da
mucoadesão e retenção de fármacos**

Dissertação apresentada ao Programa de Pós-Graduação em Química do Centro de Ciências Exatas da Universidade Federal do Espírito Santo como requisito parcial para obtenção do título de Mestre em Química, na área de Físico-Química de Soluções e Polímeros.

Orientador: Dr. Eloi Alves da Silva Filho

Coorientador: Dr. Arlan da Silva Gonçalves

Vitória
2018

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

S587e Silveira, Cleverton de Oliveira Cavalcanti da, 1985-
Estudos computacionais do modelo de mucina 2 e sua
interação com quitosanas tioladas : uma nova estratégia para o
estudo da mucoadesão e retenção de fármacos / Cleverton de
Oliveira Cavalcanti da Silveira. – 2018.
99 f. : il.

Orientador: Eloi Alves da Silva Filho.

Coorientador: Arlan da Silva Gonçalves.

Dissertação (Mestrado em Química) – Universidade Federal
do Espírito Santo, Centro de Ciências Exatas.

1. Dinâmica molecular. 2. Quitosana. 3. Docking. I. Silva
Filho, Eloi Alves da. II. Gonçalves, Arlan da Silva. III. Universidade
Federal do Espírito Santo. Centro de Ciências Exatas. IV. Título.

CDU: 54

Elaborado por Perla Rodrigues Lôbo – CRB-6 ES-527/O

“Estudos computacionais do modelo de mucina 2 e sua interação com quitosanas tioladas: uma nova estratégia para o estudo da mucoadesão e retenção de fármacos.”

Cleverton de Oliveira Cavalcanti da Silveira

Dissertação submetida ao Programa de Pós-Graduação em Química do Centro de Ciências Exatas da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do Grau de Mestre em Química.

Aprovado em 15/03/2018 por:



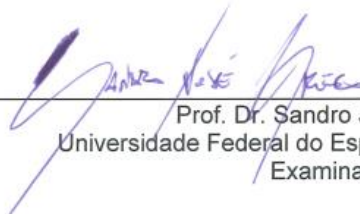
Prof. Dr. Eloi Alves da Silva Filho
Universidade Federal do Espírito Santo
Orientador



Prof. Dr. Arlan da Silva Gonçalves
Instituto Federal do Espírito Santo
Coorientador



Prof. Dr. Osmair Vital de Oliveira
Instituto Federal de São Paulo
Examinador Externo



Prof. Dr. Sandro José Greco
Universidade Federal do Espírito Santo
Examinador Interno

Universidade Federal do Espírito Santo
Vitória, março de 2018

Dedico este trabalho à minha referência em termos de força, honestidade e sensibilidade, Elisbeth de Oliveira Silveira – minha mãe.

AGRADECIMENTOS

Desejo agradecer aos meus colegas de caminhada Fabricio Uliana e Stêner Romanel Ambrozio por, sempre que solicitados, ajudarem a dirimir minhas dúvidas no trato com a máquina que, diga-se de passagem, não foram poucas. Ao professor Eloi Alves da Silva Filho pela oportunidade de trabalhar consigo, bem como pela confiança durante esses últimos dois anos. E meu agradecimento especial ao professor Arlan da Silva Gonçalves por aceitar de pronto o desafio de coorientar minha pesquisa em química computacional, sobretudo conhecendo minha relação com a máquina durante as nossas aulas de físico-química. Agradeço também a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela bolsa [concedida](#) durante todo o tempo da pesquisa.

LISTA DE FIGURAS

Figura 1. Representação estrutural do processo de desacetilação da quitina. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18.....	17
Figura 2. As cinco quitosanas modificadas mais promissoras para fins mucoadesivos. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18	19
Figura 3. Representação de dois tipos de interações, eletrostática e ligação de dissulfeto, que contribuem para a mucoadesão entre membranas mucosas e quitosanas modificadas com grupos sulfidrilas	20
Figura 4. Estrutura de quitosanas tioladas com omissão dos hidrogênios. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18.....	22
Figura 5. Representação da glicoproteína MUC2. Adaptado de BANSIL e TURNER, 2006	25
Figura 6. Alinhamentos possíveis para duas sequências distintas.....	30
Figura 7. Alinhamento múltiplo realizado pelo servidor CLUSTAL OMEGA, sequência-assinatura dos domínios C do FvW de diferentes proteínas humanas, sequências obtidas em UniProt.....	31
Figura 8. Representação de dois tipos de potenciais ligados, estiramento, deformação angular e a curva de energia potencial do tipo harmônica. Adaptado de LEWARS, 2011	35
Figura 9. Representação da variação angular que dá origem ao potencial diedral. Adaptado de LEWARS, 2011	36
Figura 10. Termos de energias potenciais ligadas para o campo de força GROMOS96 (VERLI, 2014: 175)	38
Figura 11. Parte da sequência primária da MUC2 encontrada no banco de dados de mucina a partir da qual foi construído o modelo VWD2-VWD3.....	43
Figura 12. Representação estrutural dos ligantes utilizados no presente trabalho. Imagem gerada a partir do programa AVOGADRO, versão 1.1.1 (HANWELL, CURTIS e CURTIS et al. 2012).....	46
Figura 13. Domínios modelados, em A 106 primeiros resíduos modelados pelo servidor LOMETS por Threading, em B e C resíduos de 107 a 208 e 209 a 314,	

respectivamente, por modelagem comparativa usando o RaptorX. Imagem gerada pelo programa PYMOL versão 1.2r3pre (Schrödinger, LLC)	50
Figura 14. Modelo completo VWD2-VWD3 em diferentes perspectivas, obtido pela união dos domínios, (FIGURA 13), pelo servidor AIDA, A e B representação em fita, A' e B' em superfície. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	50
Figura 15. Representação em fita do modelo com ênfase nas cistinas e cisteínas..	51
Figura 16. Ilustração dos ângulos diedros ψ e ϕ em uma cadeia tripeptídica, R1 (glicina), R2 (arginina) e R3 (alanina). Imagem gerada pelo programa GMOLDEN, versão 5.0 (SCHAFTENAAR e NOORDIK, 2000) e adaptado pelo autor	51
Figura 17. Diagrama de Ramachandran, gerado pelo PROCHECK, versão 3.5 (LASKOWSK, MACARTHUR e MOSS et al. 1993), do modelo VWD2-VWD3 obtido por modelagem comparativa e Threading	52
Figura 18. Resultado da submissão do modelo VWD2-VWD3 ao servidor Verify3D	54
Figura 19. Gráfico RMSD versus Tempo de simulação do modelo VWD2-VWD3 ...	55
Figura 20. Sobreposição da conformação inicial (verde, t = 0 ns) e conformação final (azul, t = 75 ns) do modelo VWD2-VWD3, representação em fita. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	56
Figura 21. Representação para o Fator-B de cada aminoácido do modelo VWD2-VWD3, representação (esquerda) em fita e massas (direita). Do azul para o vermelho em ordem crescente de oscilação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	56
Figura 22. Evolução das ligações hidrogênio inter e intracadeia ao longo da simulação por DM	57
Figura 23. Ilustração da ASAS. Adaptado da Wikipédia.....	58
Figura 24. ASAS ao longo da simulação por DM	58
Figura 25. Sobreposição dos 4 clusters gerados a partir dos últimos 10 ns de simulação com um raio de corte de 2.5 Å. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	59
Figura 26. Melhor pose para o ligante AC-Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	60
Figura 27. Melhor pose para o ligante AMB-Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	61

Figura 28. Melhor pose para o ligante AMN-Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	61
Figura 29. Melhor pose para o ligante ATG-Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	61
Figura 30. Melhor pose para o ligante Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	62
Figura 31. Melhor pose para o ligante TBA-Qui, obtido a partir do docking molecular. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	62
Figura 32. Mudança do ligante TBA-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	63
Figura 33. Mudança do ligante AC-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	63
Figura 34. Mudança do ligante Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	63
Figura 35. Mudança do ligante AMN-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	64
Figura 36. Mudança do ligante AMB-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	64
Figura 37. Mudança do ligante ATG-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa PYMOL, versão 1.2r3pre (Schrödinger, LLC)	64
Figura 38. $\Delta G_{(\text{ligação-MM/PBSA})}$ receptor-ligante ao longo de 15 ns de simulação	65
Figura 39. Número de ligação hidrogênio receptor-ligante em função do tempo	67
Figura 40. Pontes salinas entre amidina protonada do ligante TBA-Qui, aspartato (ASP) 157 e glutamato (GLU) 274. Imagem gerada pelo programa VMD, versão 1.9.3 (HUMPHREY, DALKE e SCHULTEN, 2012).....	67
Figura 41. Distâncias entre amidina protonada do ligante TBA-Qui, aspartato (ASP) 157 e glutamato (GLU) 274.....	68
Figura 42. Número de ligação hidrogênio receptor-ligante somente para o substituinte do hexâmero em função do tempo.....	68
Figura 43. RMSD para os ligantes em função do tempo	69
Figura 44. Comparação dos RMSF do modelo VWD2-VWD3 sem (preto) e com os diferentes ligantes (coloridos).....	69

LISTA DE TABELAS

Tabela 1. Classificação de ordem de mucoadesão por diferentes métodos	22
Tabela 2. Parâmetros de classificação para os modelos do RaptorX.....	49
Tabela 3. Resultados de <i>docking</i> para cada ligante com seu respectivo <i>cluster</i>	60
Tabela 4. Média dos resultados de $\Delta G_{(\text{ligação-MM/PBSA})}$ para cada ligante ao longo da simulação por DM, com detalhes da contribuição de cada termo para o resultado final. Ordem decrescente da esquerda para a direita.....	66

LISTA DE ABREVIATURAS E SIGLAS

AC-Qui – N,acetil-cisteína-quitosana
AMB-Qui – N,4-ácido mercaptanobenzoico-quitosana
AMN-Qui – N,6-ácido mercaptanonicotínico-quitosana
AM1-BCC – Método para geração de cargas atômicas em fase condensadas
ASAS – Área Superficial Acessível ao Solvente
ATG-Qui – N,ácido tioglicólico-quitosana
CASP – *Critical Assessment of protein Structure Prediction*
CysD – Domínios ricos em cisteínas
C – Carbono
C_α – Carbono alfa
DM – Dinâmica Molecular
FvW – Fator de von Willebrand
GDT – *Global Distance Test*
IT – Integração Termodinâmica
MC – Modelagem Comparativa
MUC2 – Mucina 2
MMPBSA – *Mecânica Molecular Poisson-Boltzmann Surface Area*
N – Nitrogênio
OPLS-AA – *Optimized Potentials for Liquid Simulations – All Atoms*
PDB – *Protein DataBank*
PEL – Perturbação da Energia Livre
PTS – Domínios de mucina ricos em Prolina Treonina e Serina
QSAR – *Quantitative Structure-Activity Relationship*
Qui – Quitosana
RMN – Ressonância Magnética Nuclear
RMSD – *Root-Mean-Square Deviation*
SCOPe – *Structural Classification of Proteins-extended*
SPR – Espectroscopia de ressonância de plasma de superfície
TBA-Qui – N,4-tiobutil-amidina-quitosana
UniProt – Banco de dados de sequências primárias de proteínas
VWD2 – Domínios D2 do fator de von Willebrand

VWD3 – Domínios D3 do fator de von Willebrand

VWD2-VWD3 – Modelo parcial de mucina 2

ΔG – Variação de energia livre de Gibbs

ΔH – Variação de entalpia

LISTA DE SÍMBOLOS

Å – Angström

fs – femtosegundos

kcal – quilocalorias

kJ – quilojoules

mol – quantidade de matéria

nm – nanômetros

nm² – área em nanômetros

ps – picossegundos

RESUMO

Quitosanas têm despertado interesse como excipiente mucoadesivo que aumenta o tempo de residência de fármacos junto às mucosas. A interação com a mucina 2 ao longo do intestino determina o grau de mucoadesão, potencializada com a inserção de substituintes tiolados em sua estrutura. Estudou-se as interações entre a glicoproteína mucina 2 e quitosanas tioladas, ranqueando-as a partir da energia livre de interação complexo receptor-ligante. A ausência de trabalhos *in silico* no estudo da mucoadesão, e possível desenvolvimento de uma metodologia para esse fim, motivou a investigação realizada a partir do desenvolvimento de um modelo de mucina 2 por modelagem comparativa e *threading*, dinâmica molecular, *docking* e MMPBSA. Quando se considerou as interações não ligadas, o ligante N-acetil-cisteína-quitosana (AC-Qui) se equiparou em termos de energia livre de ligação ao hexâmero N-tiobutil-amidina-quitosana (TBA-Qui). A quitosana (Qui) apresentou o segundo melhor $\Delta G_{(ligação)}$. Esses resultados mostraram que o grau de mucoadesão das quitosanas tioladas assumiu ordem diversa, quando se consideraram apenas as interações não ligadas, diferente dos resultados experimentais que não permitem uma especificação da natureza da interação envolvida no processo.

Palavras-chave: MUCINA 2. MUCOADESÃO. MMPBSA. DOCKING. QUITOSANA TIOLADA. DINÂMICA MOLECULAR.

ABSTRACT

Chitosans have attracted interest as a mucoadhesive excipient that increases the residence period of drugs along mucous membranes. The interaction with the mucin 2 throughout the intestine determines the level of mucoadhesion, potentiated with the insertion of thiolated substitutes in its structure. There were studies about the interactions between the glycoprotein mucin 2 and thiolated chitosans, ranking them based on the energy free of ligand-receptor complex interaction. The absence of *in silico* works in the study of mucoadhesion and the possible development of a methodology to this extent has motivated the investigation performed from the development of a mucin 2 model by the use of comparative modeling and fold recognition; molecular dynamics, docking and *mmpbsa*. When non-bonded interactions were considered, the Chitosan-N-acetyl cysteine (AC-Chi) ligand equaled itself in terms of energy free of bond to the hexamer Chitosan-thiobutylamidine (TBA-Chi) and the Unmodified Chitosan (U-Chi) displayed the second greatest $\Delta G_{(binding)}$. This result showed that the level of mucoadhesion of thiolated chitosans has assumed a diverse order, when considering only the non-binding interactions, different from the experimental results that do not allow a specification of the nature of the interaction involved in the process.

Keywords: MUCIN 2. MUCOADHESION. MMPBSA. DOCKING. THIOLATED-CHITOSAN. MOLECULAR DYNAMICS.

SUMÁRIO

1 INTRODUÇÃO	17
1.1 QUITOSANA	17
1.2 MUCOADESÃO	18
1.3 MÉTODOS PARA A DETERMINAÇÃO DA MUCOADESÃO	20
1.4 COMPARAÇÃO DO POTENCIAL MUCOADESIVO DE QUITOSANAS MODIFICADAS	21
1.5 MUCINAS E SEU PAPEL NA ADMINISTRAÇÃO DE FÁRMACOS POR VIA ORAL	24
1.6 O QUE É MODELAGEM MOLECULAR	26
1.6.1 Modelagem estrutural de proteínas	27
1.6.2 Dinâmica molecular	33
1.6.3 Docking molecular	38
1.6.4 Técnica MMPBSA para obtenção de $\Delta G_{\text{ligação}}$ entre receptor e ligante	40
2. OBJETIVOS	42
2.1 OBJETIVO GERAL	42
2.2 OBJETIVOS ESPECÍFICOS	42
3. METODOLOGIA	42
3.1 CONSTRUÇÃO DO MODELO DA MUC2 UTILIZANDO OS SERVIDORES RAPTORX e LOMETS ON-LINE	42
3.2 REFINAMENTO POR DM DO MODELO VWD2-VWD3	44
3.3 OBTENÇÃO DE DIFERENTES CONFORMAÇÕES DO MODELO VWD2-VWD3 PARA O ESTUDO DE ENSEMBLE DOCKING	45
3.4 ESTUDO POR DOCKING MOLECULAR DAS INTERAÇÕES INTERMOLECULARES DO MODELO VWD2-VWD3 COM QUITOSANAS MODIFICADAS	45
3.4.1 Desenho, otimização e estado de protonação dos ligantes	45
3.4.2 Preparação do receptor e ligante para o docking molecular	46
3.4.3 Refinamento dos resultados de docking por DM e cálculo de $\Delta G_{\text{ligação}}$ usando a técnica MMPBSA	47
3.5 RECURSOS COMPUTACIONAIS	47
4. RESULTADO E DISCUSSÃO	47
4.1 OBTENÇÃO DO MODELO MUC2	47
4.2 ANÁLISE DA DM NA ETAPA DE REFINAMENTO DO MODELO VWD2-VWD3	54
4.3 RESULTADOS DE DOCKING PARA OS HEXÂMEROS DE QUITOSANA E DIFERENTES QUITOSANAS MODIFICADAS	59

4.4 ANÁLISE DA DM DO COMPLEXO RECEPTOR-LIGANTE E $\Delta G_{\text{ligação}}$ OBTIDO PELO MÉTODO <i>MMPBSA</i>	62
5 CONCLUSÃO	70
5.1 PERSPECTIVAS.....	70
6. REFERÊNCIAS BIBLIOGRÁFICAS	71

1 INTRODUÇÃO

1.1 QUITOSANA

A quitosana (Qui) é um biopolímero natural obtido quando a N-desacetilação da quitina ultrapassa os 50%, esse processo ocorre em meio básico (NaOH concentrado) ou, alternativamente, utilizando a enzima quitina deacetilase sendo, portanto, um copolímero de 2-acetamido-2-deoxi-D-glicopiranosose e 2-amino-2-deoxi-D-glicopiranosose unidas por ligações glicosídicas β (1 \rightarrow 4). A Figura 1 mostra o processo de desacetilação da quitina representada em perspectivas de Haworth (NELSON e COX, 2014). Pela convenção adotada, se a hidroxila do carbono anomérico (C₁), envolvida nas ligações glicosídicas, nessa projeção, estiverem apontando para o mesmo lado que o C₆, têm-se as ligações do tipo β , se apontarem para lados contrários, tem-se o tipo α (NELSON e COX, 2014).

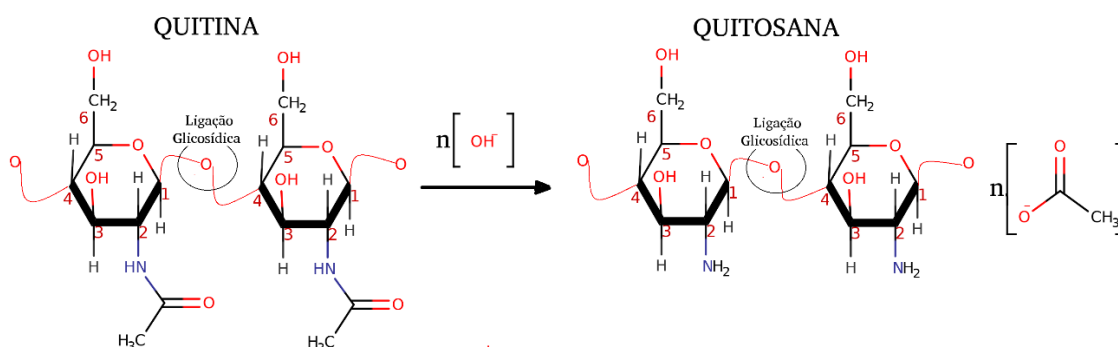


Figura 1. Representação estrutural do processo de desacetilação da quitina. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18

A Qui vem se mostrando promissora em uma série de aplicações, tais como: tratamento de água, indústria de alimentos, cosmética, farmacêutica, na agricultura e biomedicina. Essas aplicações se devem a uma série de propriedades exibidas por esse polímero, podemos citar sua viscosidade, mucoadesão, comportamento polieletrônico, capacidade de formar filmes, membranas, capacidade quelante, gelificante, entre outras (SUDHEESH, AJAY e OMOTAYO *et al.* 2013). Outro fator de grande apelo às pesquisas com Qui está ligado ao fato de ser um material proveniente de fonte renovável, biodegradável, biocompatível e não tóxico aos mamíferos.

Sua limitada solubilidade em uma faixa de pH mais ampla é um obstáculo à sua larga aplicação. Entretanto sua solubilidade e demais propriedades podem ser moduladas, ligando diversas moléculas ou polímeros por meio de inúmeras reações. Essas modificações são possíveis devido aos grupos nucleofílicos reativos presentes na Qui como a amina primária no carbono 2, a hidroxila do carbono 6 e, em menor grau, a hidroxila do carbono 3 (FIGURA 1) (SUDHEESH, AJAY e OMOTAYO *et al.* 2013). Vários trabalhos já foram realizados nesse sentido, resultando em inúmeras Qui quimicamente modificadas (KUMAR S, VIMAL e KUMAR A 2016; BADHE, NANDA e CHEJARA *et al.* 2015).

1.2 MUCOADESÃO

Dentre essas Qui modificadas, as que contêm grupos tióis têm sido cotadas como veículos de fármaco, excipientes multifuncionais que exibem várias propriedades, entre elas, uma excelente propriedade bioadesiva. Bioadesão é o estado no qual dois materiais, sendo pelo menos um de natureza biológica, são mantidos unidos por um período de tempo, quando um desses materiais é a membrana de uma mucosa, a bioadesão é definida como mucoadesão (CARVALHO, CHORILLI e GREMIÃO, 2014).

Em um artigo de revisão recente KUMAR S, VIMAL e KUMAR A (2016) elencaram as cinco quitosanas modificadas que apresentaram os melhores resultados em testes de mucoadesão, como se pode observar na Figura 2. Dessas, quatro são tioladas.

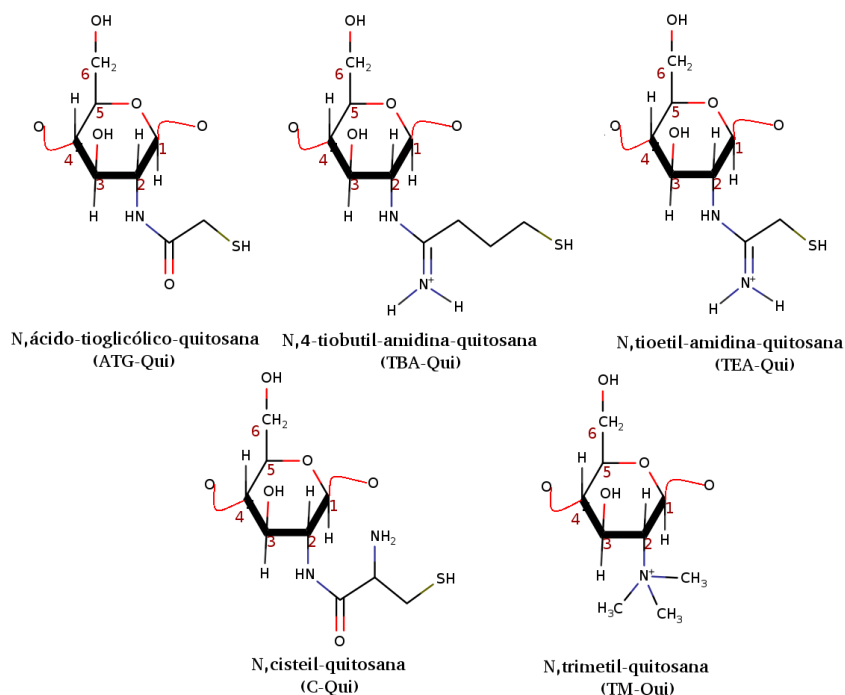


Figura 2. As cinco quitosanas modificadas mais promissoras para fins mucoadesivos. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18

A propriedade mucoadesiva da Qui se deve primeiramente às interações eletrostáticas estabelecidas entre o biopolímero catiônico, quando protonado, e as glicoproteínas aniônicas que compõem o gel formador de muco presente em todo o sistema gastrointestinal.

Sua capacidade mucoadesiva é proporcional ao grau de desacetilação e aumento da massa molecular média (KUMAR S, VIMAL e KUMAR A, 2016), podendo ser potencializada nas quitosanas tioladas por ligação dissulfeto estabelecida entre os grupos tióis com os resíduos de cisteína presente em proteínas e glicoproteínas (CHIVANGKUL, PENGPRECHA e PADUNGROS *et al.*, 2014), promovendo, assim, um maior tempo de residência da formulação administrada por via oral e, conseqüentemente, melhorando a absorção dos fármacos.

Na Figura 3, observam-se os principais tipos de interações químicas envolvidas na mucoadesão.

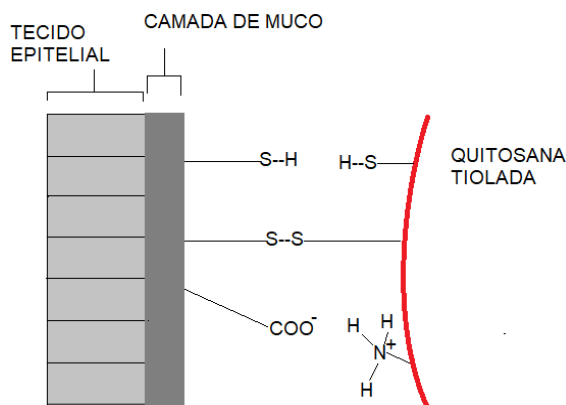


Figura 3. Representação de dois tipos de interações, eletrostática e ligação de dissulfeto, que contribuem para a mucoadesão entre membranas mucosas e quitosanas modificadas com grupos sulfidrilas

1.3 MÉTODOS PARA A DETERMINAÇÃO DA MUCOADESÃO

Existem vários métodos para a determinação da mucoadesão de polímeros, tanto *in vivo*, *in vitro* e *ex vivo*, que são aplicados no estudo dessa propriedade baseados em diferentes princípios como, por exemplo, o método baseado em medidas reológicas que se fundamenta no aparecimento de sinergismo viscosimétrico entre o polímero e a glicoproteína quando suas soluções são misturadas, fazendo com que a viscosidade da mistura seja maior que a soma da viscosidade de cada um em separado (CARVALHO, CHORILLI e GREMIÃO, 2014).

No escoamento de líquidos, uma porção da mucosa fresca de animal é fixada em um suporte e, discos do mucoadesor são, então, colocados em contato com essa mucosa, fazendo um fluxo contínuo de solução tampão com pH fisiológico percorrer o sistema. O grau de mucoadesão é proporcional ao tempo necessário para retirar por erosão todo o mucoadesor da superfície da mucosa (CARVALHO, CHORILLI e GREMIÃO, 2014).

Um dos métodos mais populares é a força de destacamento em que é utilizado um texturômetro para obter a curva da força necessária para destacar um disco do mucoadesor da superfície da mucosa animal fresca ou disco de mucina, a força de destacamento é a força máxima necessária para separar as duas superfícies. A partir desse parâmetro é possível obter outro: o trabalho de destacamento obtido pela integração da curva de força (CARVALHO, CHORILLI e GREMIÃO, 2014).

O cilindro rotatório é bem parecido com o escoamento de líquidos, a diferença consiste na fixação da mucosa na parede de um semicilindro de metal cortado longitudinalmente, imerso em uma solução tampão. A erosão dos discos de mucoadesor se dá pela força de cisalhamento gerada pelo movimento giratório do cilindro, o grau de mucoadesão é dado pelo tempo necessário para completa remoção do mucoadesor (CARVALHO, CHORILLI e GREMIÃO, 2014).

Titulação Calorimétrica, com ele se obtém a variação de entalpia (ΔH) de complexação entre a mucina e o mucoadesor. Além desses, estão sendo utilizados testes em humanos e animais, espectroscópicos tais como espectroscopia no ultravioleta-visível, espectroscopia de ressonância de plasma de superfície (SPR) (CARVALHO, CHORILLI e GREMIÃO, 2014), os quais não se abordarão aqui por fugirem ao escopo desse trabalho.

1.4 COMPARAÇÃO DO POTENCIAL MUCOADESIVO DE QUITOSANAS MODIFICADAS

Em um trabalho que buscou comparar o potencial mucoadesivo de diversas quitosanas tioladas, MUELLER, VERROKEN e IQBAL *et al.* (2011) utilizaram os métodos de força de destacamento, cilindro rotatório e reológico para ranquear seis quitosanas tioladas (FIGURA 4).

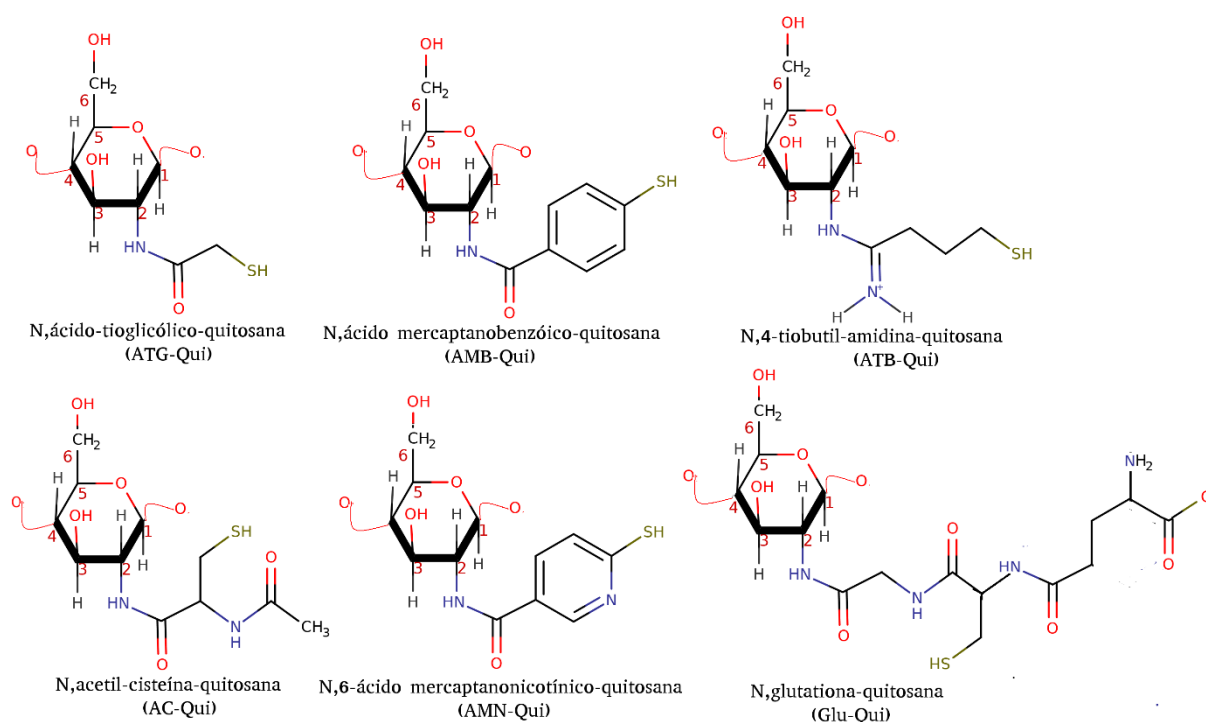


Figura 4. Estrutura de quitosanas tioladas com omissão dos hidrogênios. Imagem gerada a partir do programa MARVINSKETCH, versão 17.18

O resultado desse trabalho pode ser visto na Tabela 1. Como se pode notar a ordem de mucoadesão varia dependendo do método utilizado, apesar da não convergência de resultados entre os métodos, os autores concluem pela seguinte ordem de mucoadesão: TBA-Qui > AMB-Qui > Glu-Qui > AMN-Qui > AC-Qui > ATG-Qui > Qui.

Tabela 1. Classificação de ordem de mucoadesão por diferentes métodos

ORDEM DECRESCENTE	CILINDRO ROTATÓRIO	FORÇA DE DESTACAMENTO	REOLÓGICO
1	TBA-Qui	AMN-Qui	TBA-Qui
2	AMB-Qui	TBA-Qui	ATG-Qui
3	AMN-Qui	AMB-Qui	Glu-Qui
4	Glu-Qui	Glu-Qui	AC-Qui
5	ATG-Qui	AC-Qui	AMB-Qui
6	AC-Qui	ATG-Qui	AMN-Qui
7	Qui	Qui	Qui

O aumento na mucoadesão a partir da funcionalização da amina primária da quitosana é surpreendente, indo de 2 horas o tempo necessário para a remoção de

toda a Qui da superfície da mucosa no cilindro rotatório, até 180 horas para o TBA-Qui. Salieta-se que vários são os fatores que contribuem para formação das ligações covalentes entre grupos tióis do mucoadesor e muco, resultando em um bom desempenho, como por exemplo: o pKa do grupo sulfidrilica estar dentro da faixa de pH encontrada no intestino que pode variar de 5,5 até 7,5, pois é importante que esse grupo esteja em sua forma desprotonada, a fim de favorecer as ligações covalentes. Outro fator é a flexibilidade da cadeia do ligante o que permitiria uma interpenetração entre muco e ligante de modo a favorecer o contato também necessário para as ligações dissulfeto; que o ligante seja capaz de preservar a carga líquida positiva da cadeia polimérica como faz o TBA-Qui e o TEA-Qui referido anteriormente, além de ter um bom caráter hidrofílico (MUELLER, VERROKEN e IQBAL *et al.* 2011).

BADHE, NANDA e CHEJARA *et al.* (2015) levaram a efeito uma modificação da quitosana com polietilenoglicol e cisteína assistida por micro-ondas. Além do trabalho experimental, eles fizeram um pequeno estudo de energia de complexação *in silico* da quitosana modificada com um modelo da principal glicoproteína envolvida no fenômeno de mucoadesão intestinal, a mucina. A utilização de modelagem molecular, ainda que de modo incipiente, se mostrou uma inovação quando aplicado ao estudo de mucoadesão. Embora as metodologias para estudos *in silico* já estejam em um estágio avançado, ao que parece foi a primeira vez que se fez uso delas como um auxiliar nas pesquisas de mucoadesão de quitosanas modificadas. Um dos problemas que vem à tona é saber como obtiveram e avaliaram o modelo de mucina utilizado no trabalho, pois os autores não deixam claro no artigo.

A partir do caminho apontado por BADHE, NANDA e CHEJARA *et al.* (2015) e MUELLER, VERROKEN e IQBAL *et al.* (2011), o estudo por meio da Modelagem Molecular das interações entre quitosanas tioladas e a mucina abre perspectivas para a obtenção de um método *in silico* capaz de fornecer informações *a priori* sobre propriedades mucoadesivas de quitosanas modificadas, a fim de auxiliar no planejamento de novos polímeros mucoadesivos.

Mas, para isso, é importante passar em revista os fundamentos das técnicas que compreendem a Modelagem Molecular, bem como o que se sabe sobre a principal

glicoproteína de interesse em um estudo dessa natureza, o que será feito nas próximas subseções.

1.5 MUCINAS E SEU PAPEL NA ADMINISTRAÇÃO DE FÁRMACOS POR VIA ORAL

As mucinas são uma classe de glicoproteínas de alto peso molecular, sendo as principais responsáveis pela formação do gel que recobre as paredes das mucosas. Elas desempenham um papel de barreira física contra o ataque de ácidos, peptidases, sais biliares e micro-organismos, além da função de lubrificação. Uma representante das mucinas de extrema relevância para a administração de fármacos por via enteral é a glicoproteína multidomínio MUC2, que está predominantemente presente no intestino delgado, sua cadeia de aminoácidos tem 5179 resíduos (*UniProt*) (<http://www.uniprot.org/uniprot/Q02817>) e é composta por vários domínios tipo D e C (<http://pfam.xfam.org/family/PF00093>) do Fator de von Willebrand (FvW).

O FvW é uma glicoproteína de alto peso molecular que pode formar homopolímeros da ordem de 15 MDa, é encontrado circulando no plasma sanguíneo, responsável por mediar processos hemostáticos como adesão plaquetária ao subendotélio e promover a agregação plaquetária em locais nos quais as paredes de vasos sanguíneos tenham sido lesionadas, além de ser responsável pelo transporte do Fator de Coagulação VIII.

Erros na biossíntese do FvW levam às disfunções hemostáticas conhecidas como doença de von Willebrand (JORIEUX, FRESSINAUD e GOUEMAND *et al.* 2000). Estudos experimentais apontam que as cisteínas presentes nos domínios D e C do FvW estão todas estabelecendo ligação dissulfeto intra e intercadeia (O'LEARY, HAMILTON e DEANE *et al.* 2004, BONTHRON, HANDIN e KAUFMAN *et al.* 1986), com base nisso, pode-se especular que as cisteínas nos tipos D e C da MUC2 também estejam fazendo ligação dissulfeto (Johansson, Larsson e Hansson *et al.* 2011).

Além dos quatro domínios tipo D de FvW, sendo três na extremidade N-terminal e um na C-terminal e dois domínios tipo C de FvW na parte C-terminal, a MUC2

também apresenta dois domínios CysD, ricos em cisteína com aproximadamente 110 aminoácidos dos quais as 10 cisteínas presentes estão formando 5 ligações dissulfeto intra domínios (AMBORT, VAN DER POST e JOHANSSON, *et al.* 2011). Apresenta, também, dois domínios tipo mucinas compostos por repetições em *tandem* de prolinas, treoninas e serinas (PTS), o primeiro domínio PTS tem aproximadamente 500 resíduos, o segundo e maior domínio tem aproximadamente 2300 resíduos, esses dois domínios estão separados pelo segundo domínio CysD.

As glicosilações ocorrem basicamente nesses dois Domínios PTS, são tão numerosas que os oligossacarídeos podem corresponder 80% da massa molecular da MUC2, nessas duas regiões PTS as ligações são do tipo O-glicosiladas, os glicanos lhes conferem um aspecto alongado e de natureza hidrofílica responsável pela absorção de água para a formação do gel, além de proporcionar alta resistência às ações das proteases, principalmente impedindo que elas alcancem a cadeia peptídica da MUC2. Outro tipo de glicosilação presente, porém em menor grau, são os oligossacarídeos N-ligados, que se ligam preferencialmente em resíduos de asparagina ao longo de toda a glicoproteína (PELASEYED, BERGSTRÖM e GUSTAFSSON *et al.* 2014). A Figura 5 ilustra os vários domínios da MUC2 **a** e **b** e sua capacidade de se polimerizar **c** e **d**

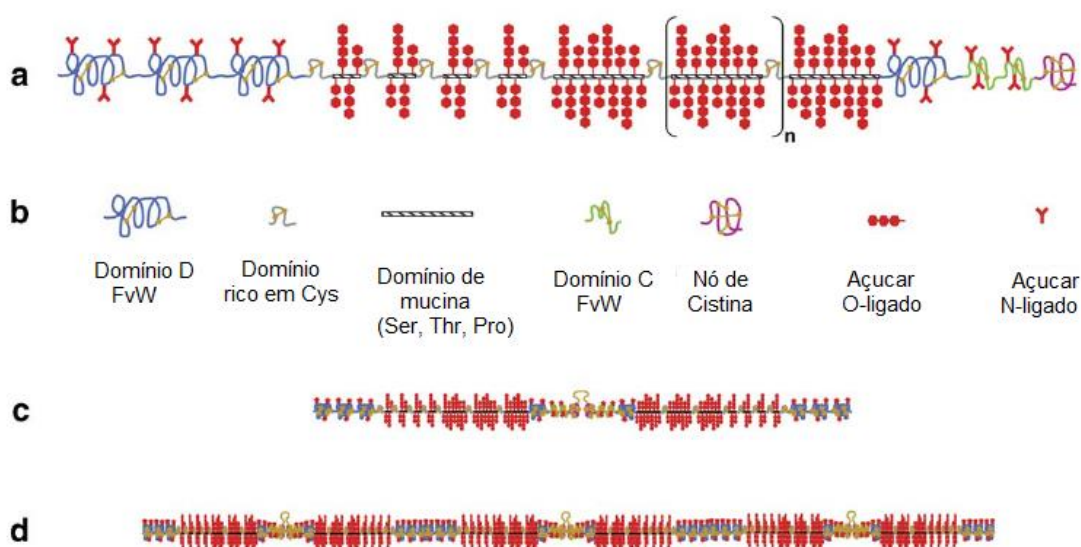


Figura 5. Representação da glicoproteína MUC2. Adaptado de BANSIL e TURNER, 2006

A MUC2 é o principal elemento de interação com a quitosana tiolada no intestino, dessa forma, a compreensão mais detalhada das interações entre os grupos tióis da

glicoproteína e do polímero passa pelo desenvolvimento de um modelo de MUC2 representativo do sistema biológico, uma vez que não há disponíveis modelos experimentalmente obtidos.

Entretanto, a utilização de metodologias *in silico*, apesar de proporcionar vantagens imensuráveis, esbarra no que se costuma chamar de custo computacional, por exemplo, para escolha da DM como ferramenta para a investigação de sistemas macroscópicos é necessário levar-se em consideração alguns pontos como o número de partículas do sistema simulado: um modelo de proteína muito grande implica em uma caixa grande que precisa ser preenchida com solvente e, quanto maior a caixa, mais moléculas de solvente serão necessárias, impondo ao computador realizar maior número de integrações para determinar a evolução do sistema, tornando a obtenção das trajetórias mais demorada.

Além disso, o sistema simulado precisa ser representativo do sistema real (MORGON; COUTINHO, 2007 p. 414). Um bom resultado passa pela escolha de uma estratégia que permita minimizar o tamanho do sistema e maximizar a sua representatividade.

1.6 O QUE É MODELAGEM MOLECULAR

Para SCHLICK (2006: 3), a modelagem molecular “[...] é a ciência e a arte que estuda a estrutura e função moleculares através da construção e cálculos de modelos”. A construção de modelos pode ir desde a utilização de bolas e varetas de plástico, até a última tecnologia na criação de imagens. Já os cálculos abrangem *ab initio*, semi-empírico, dinâmica molecular (DM), Monte Carlo, relação estrutura/atividade (QSAR), *docking* molecular, entre outros.

O elemento arte da definição vem de uma inspiração na qual diz que “A arte é uma mentira que nos ajuda dizer a verdade” (Picasso *apud* SCHLICK, 2006: 3), pois os modelos são uma simplificação de complexos fenômenos, no entanto, essa simplificação pode nos ajudar a enxergar detalhes que de outro modo seria impraticável. Na sequência, abordaram-se várias técnicas que compreendem a modelagem molecular.

1.6.1 Modelagem estrutural de proteínas

As estruturas de proteínas, em princípio, podem ser conhecidas utilizando as técnicas de difração de raios-X e ressonância magnética nuclear (RMN), porém nem sempre é possível obter a estrutura de uma proteína utilizando essas técnicas. Por exemplo, quando uma proteína não cristaliza, a despeito de inúmeras tentativas. Em outros casos, cristaliza de modo desordenado, dificultando o processo de obtenção da estrutura terciária. Pode ocorrer, também, o caso de uma determinada proteína não ter uma organização estrutural bem definida, pois um terço do proteoma humano é composto de proteínas com estruturas naturalmente desordenadas (NELSON e COX, 2014).

A utilização de RMN fica limitada a proteínas abaixo de aproximadamente 35 kDa, inviabilizando sua utilização em proteínas muito grandes, pois, para proteínas maiores, as velocidades de decaimento são muito lentas e a ressonância de hidrogênio torna-se larga e, portanto, difícil de ser observada (SHRIVER, ATKINS e OVERTON *et al.* 2008: 741, SCHLICK, 2006: 3).

Apontadas as limitações dessas técnicas, soma-se a quantidade imensa de sequências primárias que estão sendo depositadas em bancos de dados como o *Universal Protein Resource* (UniProt) (<http://www.uniprot.org/>), resultado de pesquisas em genômica e proteômica. Dessa forma, é natural e necessário compreender como essas novas proteínas descobertas atuam no sistema biológico, para tanto é preciso saber como estão organizadas no espaço, ou seja, é imprescindível ter acesso a sua conformação 3D, pois a sua função biológica é dependente dessa organização espacial. A biologia computacional (SCHLICK, 2006: 33-34) ou bioinformática (NELSON e COX, 2014: 105) é um segmento da modelagem molecular que se ocupa desses problemas.

Existem basicamente três técnicas de modelagem molecular de proteínas, *ab initio*, modelagem comparativa (MC) e *Threading*. A técnica *ab initio* consiste na modelagem de proteínas com base em princípios físicos de energia, contudo, só é possível aplicá-la a proteínas com no máximo 200 aminoácidos, neste trabalho utilizou-se apenas as técnicas de MC e *Threading*, por essa razão não se

aprofundou na abordagem da técnica *ab initio*.

1.6.1.1 Modelagem estrutural de proteínas por MC

Duas proteínas são ditas homólogas se suas sequências de aminoácidos e funções são semelhantes, nesse caso elas podem ser consideradas integrantes da mesma família proteica e, muito provavelmente tenham evoluído a partir de um ancestral comum (NELSON e COX, 2014: 106). Como dito anteriormente, a função de uma proteína depende sobremaneira de sua conformação espacial específica. Alteração nessa conformação decorrente de pequenas mudanças na sequência de aminoácidos pode levar, em muitos casos, a uma perda de função ou função deletéria para a célula.

Nesses casos, no entanto, a seleção natural atua e o organismo portador dessas alterações fica para trás no caminho evolutivo (ALBERTS, JOHNSON e LEWIS *et al.* 2009: 137). Por essas razões, existem mecanismos para evitar modificações de aminoácidos em posições chave da estrutura proteica, seja em sítios ativos, ou em posições que proporcionem estabilidade estrutural do estado nativo. Como consequência, observa-se que a estrutura terciária é mais bem conservada que a estrutura primária, de forma que se possa estabelecer, por exemplo, que 40% de identidade de sequência implica em mais de 90% de sobreposição da estrutura terciária com 3,5 Å de precisão quando se considera os C α . Se a semelhança de sequência sobe para 50% a sobreposição é completa (SCHLICK, 2006: 35-36).

Com base nisso, desenvolveu-se métodos capazes de comparar duas sequências de proteínas ou uma sequência com outras, cujas estruturas já sejam conhecidas e depositadas em banco de dados como o PDB (*Protein Data-Bank*) (<http://www.rcsb.org/pdb/home/home.do>). As chances de se encontrar uma estrutura no banco de dados relacionada com uma sequência escolhida aleatoriamente variam de 20% a 70% (SAXENA e SANGWAN, 2013) e as estruturas que apresentem boas identidades de sequência são utilizadas como moldes (*templates*) para a construção do modelo de interesse (*target*), nesse processo incluem identificar regiões estruturalmente conservadas e prever as regiões variáveis, modelar as cadeias laterais, refinar e avaliar os modelos (GINALSKI, 2006).

As relações evolutivas entre duas proteínas são, portanto, identificadas a partir da semelhança entre suas sequências, dessa forma a etapa de alinhamento de sequências é uma etapa essencial para os métodos tanto de MC como de *Threading*.

Ao longo do processo evolucionário, mutações no DNA podem levar a substituição, depleção ou adição de aminoácidos nas sequências proteicas, de modo que um alinhamento global pode não permitir encontrar homologia entre duas sequências. Dessa forma, os programas disponíveis podem encontrar bons alinhamentos em alguns segmentos, ao passo que em outros não. Para resolver esse problema, eles inserem *gaps* nas sequências, de modo a obterem bons alinhamentos entre segmentos. Porém a cada *gap* imposto, há uma penalidade na pontuação global do alinhamento, estabelecem-se penalidades mais brandas quando diferenças conservativas são detectadas.

A título de exemplo, a substituição em uma determinada posição de glutamato por aspartato, ambos com cargas negativas. Contudo, diferente de uma substituição, por uma fenilalanina (NELSON e COX, 2014: 106).

Duas sequências podem ser alinhadas de diferentes formas, por exemplo, na Figura 6 abaixo, tem-se duas sequências de polipeptídeo X e Y, para cada alinhamento é dado uma pontuação, somando-se todas as penalidades e apresentando um resultado final que ranqueia os vários alinhamentos possíveis.



Figura 6. Alinhamentos possíveis para duas sequências distintas

Os programas e servidores *web* destinados a alinhamentos de sequências podem, dependendo de seus algoritmos, realizar alinhamentos aos pares, ou seja, alinhar a sequência alvo com cada uma das sequências contidas nos bancos de dados. Porém, estudos mostram que o alinhamento aos pares é capaz de encontrar metade das relações evolutivas existentes, quando se considera uma relação de 20% a 30% de identidade de sequência.

Entretanto, os algoritmos de alinhamentos múltiplos, desdobramento dos alinhamentos em pares, quando a sequência alvo é confrontada com múltiplas sequências dos bancos de dados ao mesmo tempo, é possível melhorar em até três vezes a capacidade de encontrar relação evolutiva quando a identidade de sequência fica abaixo de 30%. Esse tipo de alinhamento permite encontrar, também, o que se chama de sequência-assinatura, aminoácidos em posições específicas coincidentes em membros de uma mesma família e que estão diretamente ligados ao desempenho de uma determinada função para aquela família de proteínas (NELSON e COX, 2014: 107).

Na Figura 7, traz-se um exemplo de alinhamento múltiplo, realizado pelo servidor *CLUSTAL OMEGA* (SIEVERS, WILM e DINEEN *et al.* 2011), estas sequências

pertencem aos domínios C do FvW de diferentes proteínas multidomínios como a mucina 2 (MUC2), Chordin, FvW. Os domínios C do FvW são caracterizados por serem ricos em cisteínas e apresentarem a seguinte sequência-assinatura $C_2XXC_3XC_4$ e $C_8C_9XXC_{10}$, os X podem ser quaisquer outros aminoácidos. O domínio C de FvW da Chordin humana já teve sua estrutura resolvida experimentalmente e apresenta todas as suas cisteínas na forma de cistinas (O'LEARY, HAMILTON e DEANE *et al.* 2004).



Figura 7. Alinhamento múltiplo realizado pelo servidor CLUSTAL OMEGA, sequência-assinatura dos domínios C do FvW de diferentes proteínas humanas, seqüências obtidas em UniProt

As estruturas de proteínas obtidas a partir da MC podem ser utilizadas para estudos de diversas naturezas, desde interação proteína/ligante, proteína/proteína até correlações fenotípica e genotípica.

A precisão das informações obtidas com esses modelos vai depender da proximidade evolutiva da estrutura no banco de dados e o alvo desejado. Modelos obtidos por MC que compartilham de no mínimo 50% de identidade de seqüência apresentam variação no Desvio da Raiz Média Quadrática com a sigla em inglês *RMSD* (*Root-Mean-Square Deviation*) de aproximadamente 1Å, quando se compararam os Cα da estrutura experimental que serviu de molde e o modelo obtido. De 30% a 50% de identidade de seqüência, molde e modelo coincidem em 80% em suas estruturas, nos eventos *CASP* (*Critical Assessment of Techniques for Protein Structure Prediction*) que avaliam a qualidade e avanços das técnicas de modelagem estrutural de proteínas, a variação do *RMSD* fica entre 2 e 3 Å com os

erros principalmente localizados nas alças da proteína alvo. As semelhanças entre as estruturas podem cair para 55% quando a identidade de sequência fica entre 20% e 30%, abaixo de 20% metade dos resíduos ficam desalinhados (GINALSKI, 2006).

1.6.1.2 Modelagem estrutural de proteínas por *Threading*

A técnica de modelagem por *Threading* é baseada no padrão de enovelamento de estruturas proteicas, já elucidadas, permitindo a modelagem de proteínas de grande porte, mesmo aquelas que não tenham estruturas homólogas depositadas nos bancos de dados de estruturas terciárias.

Essa técnica também se fundamenta no princípio, já mencionado, de que a estrutura terciária é bem mais conservada que a sequência primária, porém, ainda mais restritiva quando se considera a estrutura secundária.

Os padrões de dobramento identificados e mantidos por banco de dados como *SCOPe* (<http://scop.berkeley.edu/statistics/ver=2.06>) (FOX, BRENNER e CHANDONIA, *et al.* 2015) não passam de 1300 dobras distintas, 1221 para ser exato, essas dobras são catalogadas a medida que as sequências das proteínas vão tendo suas estruturas determinadas. A taxa de identificação de novas dobras tem diminuído bastante, indicando que se está chegando próximo da exaustão.

Para se estabelecer um parâmetro, em 2009 o número de estruturas terciárias no *PBD* era de 61814 e o número de dobras era 1195, hoje o número de estruturas terciárias depositadas mais que dobrou, porém, o número de surgimento de novas dobras não acompanhou o mesmo ritmo.

Threading, ou reconhecimento de dobra, busca determinar o padrão de dobra de uma sequência, cuja estrutura seja desconhecida, comparando com dobras depositadas em bancos de dados, e selecionando aquelas que tiverem um melhor ajuste. Só que a ênfase agora é dada a estrutura secundária. Existem basicamente duas abordagens para essa metodologia, uma é baseada no critério de energia, ou seja, a sequência problema é alinhada com cada uma das dobras presentes na

biblioteca de dobras e, selecionado aquela dobra a partir da qual apresentou uma menor energia, em seguida são considerados os termos de energia de solvatação e hidrofóbicas, os modelos então são ranqueados, ocupando o topo da lista aquele que apresentar menor energia final (TEGGE, WANG, CHENG, 2010, XIONG, 2006: 223-225).

Além do critério de energia, existe a abordagem baseada em perfil, em que se faz a superposição da sequência problema com as estruturas secundárias do banco de dados e, a partir disso, são geradas informações estatísticas chamadas de perfis. Esses estabelecem uma pontuação para cada aminoácido da sequência problema, após sobrepô-lo a estrutura de referência. Além disso, estabelece-se uma pontuação para exposição ao solvente, polaridade e hidrofobicidade para cada aminoácido.

A sequência problema, ou seja, a sequência da proteína cuja estrutura será determinada é, então, modelada quanto sua possível estrutura secundária, acessibilidade ao solvente e polaridade. Essas informações são comparadas com os perfis criados e aquele que melhor se adequar com a sequência problema é selecionado como *template* (TEGGE, WANG, CHENG, 2010, XIONG, 2006: 223-225).

1.6.2 Dinâmica molecular

A DM é uma técnica que compõe a classe denominada Mecânica Molecular e se fundamenta na resolução da equação de movimento de Newton para cada átomo que compõe o sistema (EQUAÇÃO 1). A DM permite acompanhar a evolução dos átomos ao longo do tempo, conhecida como trajetória, a partir da trajetória do sistema é possível extrair várias propriedades.

$$F_{xi} = \frac{\partial}{\partial t} \left(\frac{\partial x_i}{\partial t} \right) \cdot m_i = \frac{\Delta v_i}{\Delta t} \cdot m_i = a_i \cdot m_i \quad (1)$$

Nessa equação, m_i e a_i representam a massa e a aceleração de cada átomo do sistema, respectivamente. O termo F_{xi} da equação é fornecido pelos parâmetros de Campo de Forças e $\frac{\partial x_i}{\partial t}$ representa a velocidade para cada átomo, podendo ser obtida de diferentes maneiras, por exemplo, da distribuição de Maxwell-Boltzmann

partir da temperatura especificada de acordo com a Equação 2

$$\frac{3}{2} Nk_B T = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \quad (2)$$

Em que N é o número de partículas (MORGON e COUTINHO, 2007 p. 429).

Porém o formalismo newtoniano, como expresso na Equação 1, não permite um tratamento mais geral dos sistemas físicos. Nesse sentido, opta-se pela descrição lagrangiana (EQUAÇÃO 3) e hamiltoniana (EQUAÇÃO 4) da equação de movimento.

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0, \quad i = 1, 2, \dots, N \text{ equações de segunda ordem} \quad (3)$$

Sendo a função $\mathcal{L} = E_c - U$, onde E_c é a energia cinética do sistema e U sua energia potencial e, q_i e \dot{q}_i são as coordenadas e velocidades generalizadas, respectivamente.

$$\frac{\partial H}{\partial p_k} = \dot{q}_k \quad \text{e} \quad \frac{\partial H}{\partial q_k} = -\dot{p}_k, \quad k = 1, 2, \dots, 3N \quad (4)$$

Em que $H = E_c + U$ é a energia total do sistema e, \dot{p}_k e \dot{q}_k são as primeiras derivadas dos momentos e coordenadas generalizadas, respectivamente.

As equações de Hamilton se tornam uma representação mais favorável em relação as de Lagrange, pois como se pode notar são equações de primeira ordem, ao passo que as de Lagrange são de segunda ordem.

Essas equações só podem ser integradas analiticamente para casos específicos, sendo necessário soluções numéricas. Os algoritmos de integração (integradores) precisam ter algumas características que são inerentes as equações de movimento referidas acima, como reversibilidade no tempo, pois algoritmos temporalmente reversíveis conservam o volume no espaço de fases, acurácia das trajetórias e garantia de conservação da energia e outras quantidades conservadas para

simulações longas (RINO e da COSTA, 2013, p. 20-21).

Numa coleção de átomos componentes de um sistema, aplica-se, pois, as equações newtonianas, além das posições desses átomos, são necessárias também, determinar as forças as quais estão submetidos, essas forças são derivadas dos potenciais harmônicos clássicos, esse conjunto de potenciais são os Campo de Forças.

Campo de Forças é definido como um conjunto de equações (funcionais) que descrevem as energias potenciais a que cada átomo do sistema está submetido. Os potenciais utilizados são basicamente quatro.

Potenciais gerados pelos estiramentos e contrações das ligações covalentes que são compreendidas como sendo composta por molas, bem como o gerado pelas deformações angulares (FIGURA 8)

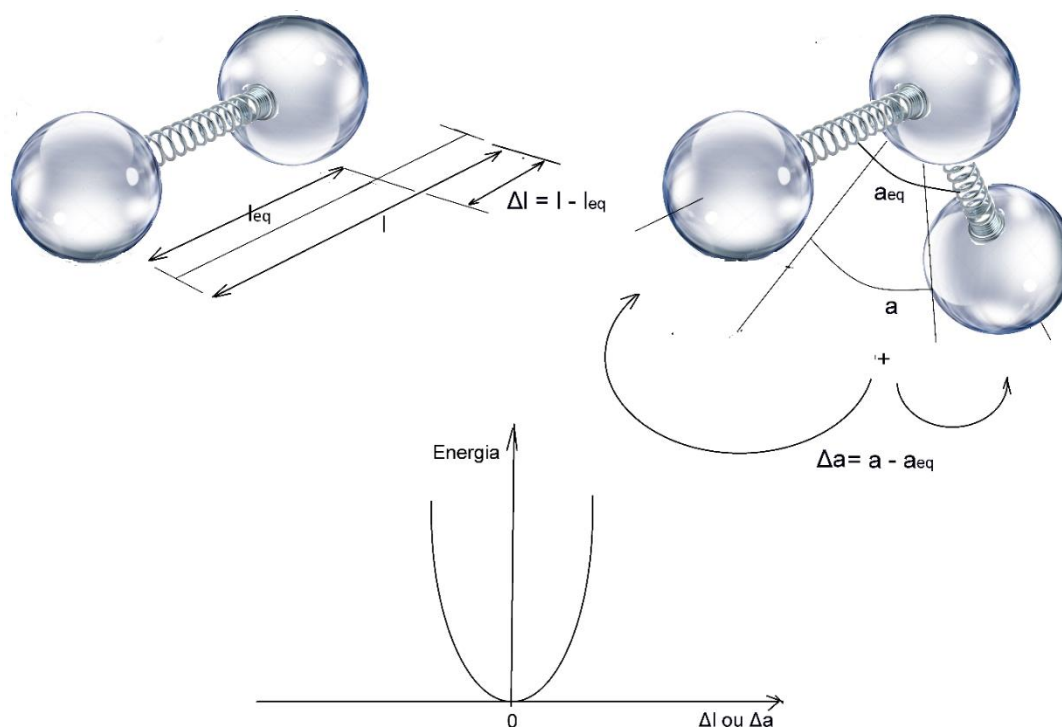


Figura 8. Representação de dois tipos de potenciais ligados, estiramento, deformação angular e a curva de energia potencial do tipo harmônica. Adaptado de LEWARS, 2011

As equações 5 e 6 descrevem os potenciais da Figura 8

$$E_{est} = K_{est}(\Delta l)^2 \quad (5)$$

$$E_{ang} = K_{ang}(\Delta \alpha)^2 \quad (6)$$

em que E_{est} e E_{ang} são os potenciais; K_{est} e K_{ang} constantes de força, Δl e $\Delta \alpha$ variação de comprimento e deformação angular, para as Equações 5 e 6 respectivamente.

Outro potencial importante é o torcional ou diedral no qual é considerado quatro átomos em sequência A-B-C-D (FIGURA 9)

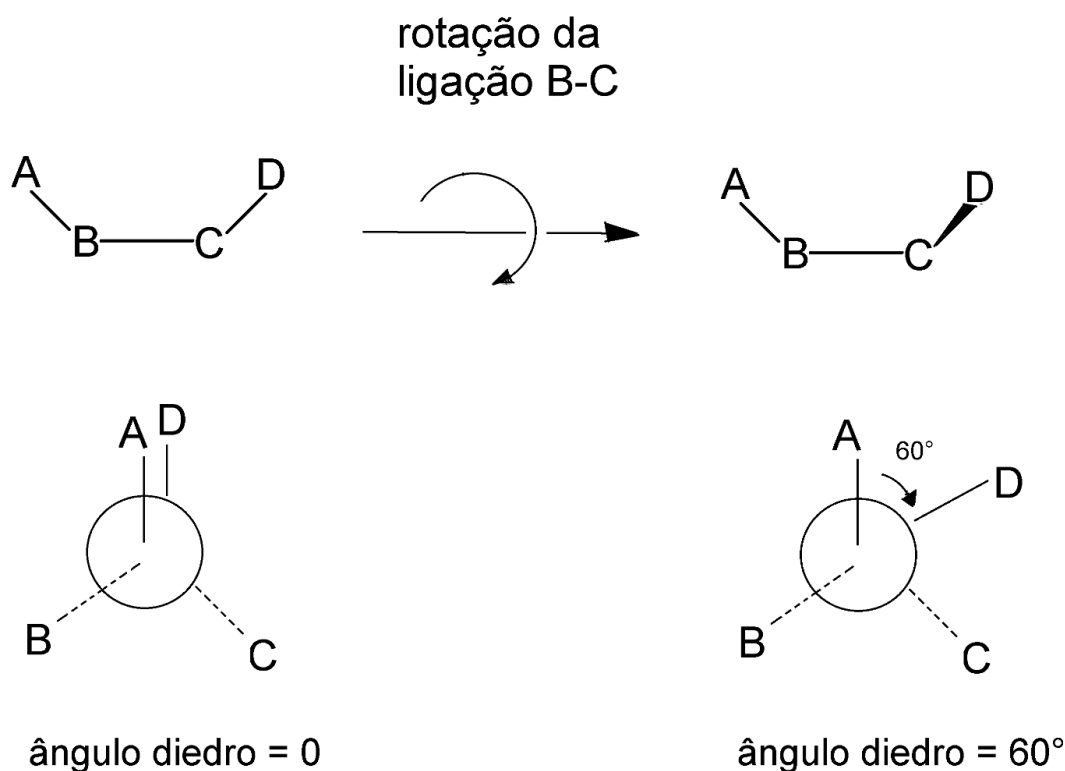


Figura 9. Representação da variação angular que dá origem ao potencial diedral. Adaptado de LEWARS, 2011

Esse potencial é descrito pela Equação 7

$$E_{died} = K_0 + \sum_{r=1}^n K_r [1 + \cos(r\theta)] \quad (7)$$

em que E_{died} é o potencial diedral e θ o ângulo diedral.

Além das equações para as energias potenciais ligadas (EQUAÇÕES 5, 6 e 7), têm-se ainda as equações que descrevem as energias potenciais não ligadas que são a energia potencial eletrostática, ou coulômbica, e de Van der Waals, ou Lennard-Jones 12-6, Equações 8 e 9, respectivamente.

$$E_{Coul} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (8)$$

$$E_{Lenn} = K_{Lenn} \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (9)$$

em que r é a distância entre os centros das partículas consideradas e σ é a distância entre as partículas na qual o potencial $E_{Lenn} = 0$.

A soma de todos esses potenciais configura o Campo de Forças (EQUAÇÃO 10)

$$E = \sum E_{est} + \sum E_{ang} + \sum E_{died} + \sum E_{Coul} + \sum E_{Lenn} \quad (10)$$

As variações nos funcionais, os diversos parâmetros obtidos para as constantes de força e as distâncias de equilíbrio e ângulos de equilíbrio dão origem aos diversos Campo de Forças existentes (LEWARS, 2011).

A Figura 10 traz um exemplo das equações que descrevem as energias potenciais ligadas relacionadas aos movimentos de estiramento de ligação, deformação angular, diedrais próprios e impróprios.

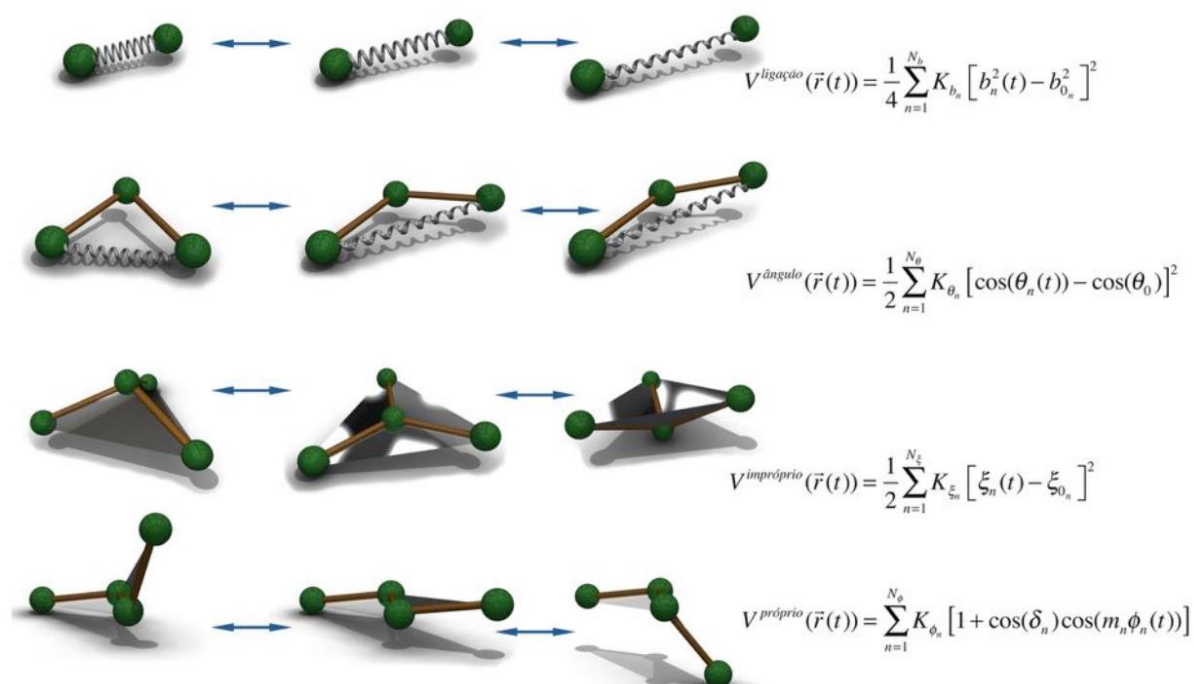


Figura 10. Termos de energias potenciais ligadas para o campo de força GROMOS96 (VERLI, 2014: 175)

Os termos que compõem essas equações são obtidos experimentalmente ou por cálculos *ab initio*. (VERLI, 2014: 183-185).

Existe uma variedade de programas capazes de realizar os cálculos de trajetória de átomos, tanto pagos quanto gratuitos, dentre eles, podemos citar LAMMPS, GROMACS, CHARMM e AMBER, os dois primeiros gratuitos (VERLI, 2014: 183-185).

1.6.3 Docking molecular

O *docking* molecular consiste na técnica computacional que visa prever os modos de interação de uma molécula chamada de ligante com um receptor e, principalmente, em quantificar em termos de energia livre de Gibbs ($\Delta G_{ligação}$) a afinidade dessa interação.

Uma alta afinidade na interação do ligante com o sítio ativo do receptor é um desdobramento de: (i) alta complementaridade estérea entre o sítio ativo do receptor

e o ligante; (ii) complementaridade de propriedades associadas as interações não ligadas, como complementaridade eletrostática e hidrofóbicas; (iii) ausência de interações repulsivas; (iv) o ligante se encontra em uma conformação energeticamente favorável (MORGON e COUTINHO, 2007 p. 489). A equação que representa a interação receptor-ligante que pode ser identificada, em termos amplos, como uma reação, exemplificado pela Equação 11.



Sendo o equilíbrio dado pela Equação 12, em que K_f é a constante de formação do complexo receptor-ligante e os [] representando a concentração em (mol/litro) das espécies em solução.

$$K_f = \frac{[Receptor-Ligante]}{[Receptor][Ligante]} \quad (12)$$

Para que a formação do complexo seja favorável sua K_f precisa ser maior que 1.

A formação do complexo receptor-ligante é dirigida por fatores entálpicos e entrópicos que podem ser estimados pelo $\Delta G_{ligação}$ por meio da Equação 13

$$\Delta G_{ligação} = \Delta H - T\Delta S = -RT \ln K_f \quad (13)$$

em que ΔH é a variação de entalpia, ΔS variação de entropia, T a temperatura do sistema, R a constante universal dos gases. Como se pode notar pela Equação 13, um processo de interação favorável deve ter a contribuição de efeitos entálpicos e entrópicos, porém, no caso de uma diminuição de entropia do sistema, o que é desfavorável, em tese, para a formação do complexo, deve ser compensado por um efeito entálpico, bem como um aumento da entalpia precisa ser compensado por um consequente aumento na entropia do sistema (VERLI, 2014: 202; MORGON e COUTINHO, 2007: 495).

Essa metodologia está fundamentada no modelo chave-fechadura, sugerido por Emil Fischer em 1894. Nesse modelo, o receptor é considerado a fechadura e o

ligante a chave e o ato de abrir a fechadura ilustraria a interação estável seguida de um efeito biológico.

Porém essa forma de ver a interação receptor-ligante não é fiel ao que sabidamente acontece durante essas interações, pois no meio biológico a flexibilidade do receptor e ligante são fatores importantíssimos para a formação do complexo. Nos programas de *docking* disponíveis atualmente os ligantes apresentam graus de liberdades translacionais, rotacionais e diedrais. Porém para o caso dos receptores com quantidades consideráveis de átomos esses graus de liberdade não são contemplados em face do custo computacional que isso demandaria. Uma visão mais adequada do processo de interação receptor-ligante é a associação dos conceitos de encaixe induzido e estados conformacionais.

No encaixe induzido, tanto receptor quanto ligante se ajustam um ao outro durante o reconhecimento molecular. Já o entendimento de estados conformacionais compreende um conjunto de conformações, adotado pelo receptor, com estruturas semelhantes e energeticamente similares. Durante a interação com o ligante, uma dessas conformações é selecionada e desloca o equilíbrio químico em sua direção, aumentando, portanto, sua probabilidade de ocorrência (VERLI, 2014: 190).

O *docking* molecular é muito eficiente para explorar o espectro conformacional de um determinado ligante em face do receptor de interesse, porém, quando se trata da previsão do $\Delta G_{\text{ligação}}$, em alguns casos pode deixar a desejar, ao ponto de não conseguirem, por exemplo, distinguir entre ligantes que apresentem diferenças de $\Delta G_{\text{ligação}}$ menor que 6 kJ/mol, comprometendo sua utilização (GENHEDEN e RYDE, 2015).

1.6.4 Técnica MMPBSA para obtenção de $\Delta G_{\text{ligação}}$ entre receptor e ligante

Além das funções de avaliação empregada pela técnica de *docking* molecular para determinar o $\Delta G_{\text{ligação}}$, são vários os métodos utilizados para estimar o $\Delta G_{\text{ligação}}$ receptor-ligante como, por exemplo, os métodos de perturbação da energia livre (PEL) e a integração termodinâmica (IT), esses são bem precisos em suas

estimativas, com erros que giram em torno de 4,18 kJ/mol, porém, demandam alto custo computacional, além de serem restritivos quando se pretende abordar sistemas com muitos átomos (MORGON e COUTINHO, 2007: 496).

A técnica *MMPBSA* apresenta uma boa acurácia e um relativo baixo custo computacional, permitindo abordar sistemas com muitos átomos e, por isso, é considerada intermediária. A parte *MM* do nome vem de mecânica molecular, pois a função para o cálculo de $\Delta G_{\text{ligação}}$, nessa técnica, contém um termo de energia ligada (E_L), composta pelos potenciais angulares, de ligação e diedrais, bem como, potenciais não ligados (E_{NL}) compostos pelas energias eletrostáticas (E_{el}) e de Van der Waals (E_{vdW}). As energias mecânico-moleculares são obtidas no vácuo. O *PB* provém da utilização da equação de *Poisson-Boltzmann* para determinar a energia livre de interação polar com o solvente (G_{polar}) e *SA* vem da área superficial acessível ao solvente, conceito utilizado para a estimativa da energia livre de interação apolar com o solvente relativo às interações apolares (G_{apolar}). A técnica *MMPBSA* utiliza a seguinte função para o cálculo de $\Delta G_{\text{ligação}}$ (EQUAÇÃO 14).

$$\Delta G_{\text{ligação}} = G_{\text{receptor-ligante}} - (G_{\text{receptor}} + G_{\text{ligante}}) \quad (14)$$

Como mencionado antes, o G para cada um dos termos da Equação 14 compreende a soma das contribuições de $E_{L,NL}$, G_{polar} e G_{apolar} (PAISSONI, SPILIOPOULOS e MUSCO *et al.* 2014), como podemos ver na Equação 15.

$$G = E_L + E_{NL} + G_{\text{polar}} + G_{\text{apolar}} \quad (15)$$

O $\Delta G_{\text{ligação}}$ *MMPBSA* é calculado a partir das trajetórias de dinâmicas individuais do receptor, ligante e receptor-ligante. Porém essa abordagem, além de ser mais custosa, pois, necessita-se realizar três dinâmicas diferentes contendo ligante, receptor e receptor-ligante, não tem apresentado resultados mais satisfatórios que aqueles obtidos com a utilização da trajetória contendo receptor-ligante apenas, nesse último caso, a E_L é cancelada na Equação 15. Embora as simulações sejam levadas a cabo com solvatação explícita, na hora do cálculo por *MMPBSA* é removido o solvente e considerado o solvente implícito para o cálculo *PBSA* (GENHEDEN e RYDE, 2015).

2. OBJETIVOS

Diante da percepção de que uma maior contribuição ao desenvolvimento das pesquisas que visem obter novos polímeros carreadores de fármaco com propriedades mucoadesivas possa ser oferecida mediante o desenvolvimento de uma metodologia *in silico* que passe pela obtenção de um modelo da glicoproteína MUC2, propõe-se como objetivos geral e específicos, os itens 2.1 e 2.2.

2.1 OBJETIVO GERAL

Propor, através da modelagem molecular, um modelo tridimensional da glicoproteína humana MUC2 que reproduza propriedades físico-químicas disponíveis na literatura, referentes às interações mucina / mucoadesores

2.2 OBJETIVOS ESPECÍFICOS

- Desenvolver modelo de MUC2 usando técnicas de modelagem estrutural de proteínas;
- Refinar o modelo gerado utilizando a técnica de DM;
- Determinar as energias de complexação do modelo de MUC2 com diversos polímeros com propriedades mucoadesivas a partir do estudo de *docking* molecular e DM;

3. METODOLOGIA

3.1 CONSTRUÇÃO DO MODELO DA MUC2 UTILIZANDO OS SERVIDORES *RAPTORX* e *LOMETS ON-LINE*

A sequência de 314 aminoácidos escolhida para compor o modelo de MUC2 foi dos resíduos de número 545 a 859 (FIGURA 11), que foram obtidos no banco de dados de mucinas, encontrado em (<http://www.medkem.gu.se/mucinbiology/databases/>) (ALBERTS, JOHNSON e LEWIS *et al.*, 2009), podendo ser obtido também, no banco de dados de sequências *UniProt* (<http://www.uniprot.org/uniprot/Q02817>), porém com pequenas diferenças na classificação dos domínios.

VEATGAGFANTWKAQSTCHDKLDWLDDPCSLNIESANYAEHWC
 SLLKKTETPFGRCHSAVDPAEYYKRCKYDTCNCQNNEDCLCAAL
 SSYARACTAKGVMLWGWREHVCNKDVGSCPNVFLYNLTTCQ
 QTCRSLSEADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCS
 CYHRGLYLEAGDVVVRQEERCVCRDGRLHCRQIRLIGQSCTAPKI
 HMDCSNLTALATSKPRALSCQTLAAGYYHTECVSGCVCPDGLM
 DDGRGGCVVEKECPCVHNNDLYSSGAKIKVDCNTCTCKRGRWV
 CTQAVCHGT

Figura 11. Parte da sequência primária da MUC2 encontrada no banco de dados de mucina a partir da qual foi construído o modelo VWD2-VWD3

A opção por utilizar essa seção da sequência de aminoácidos da MUC2 se deveu a necessidade da existência de cisteínas livres dentro da estrutura terciária do modelo para estudo. Como mencionado no item 1.5, os domínios identificados mantêm cistinas. Portanto, a escolha desse segmento (FIGURA 11) se justificou pela expectativa de que tais cisteínas estejam disponíveis, uma vez que não se encontrou estudos afirmando o contrário a respeito delas nessa região.

Como a sequência está localizada na região entre os domínios D2 de von Willebrand (VWD2) e Domínios D3 de von Willebrand (VWD3) e cujos domínios ainda não foram identificados, optou-se por chamá-lo de modelo VWD2-VWD3.

Inicialmente, copiou-se o arquivo fasta com a sequência de 314 aminoácidos na caixa disposta no site do servidor *RaptorX* (KÄLLBERG e WANG, H. e WANG, S. *et al.* 2012). Posteriormente, o servidor retornou um modelo contendo três domínios, os dois últimos com identidades de sequência aceitáveis para a técnica de modelagem comparativa. Porém o primeiro domínio composto pelos 106 primeiros resíduos não apresentou identidade de sequência minimamente adequada, com o *template*. A partir disso, lançou-se mão da modelagem *Threading*, para esse fim, utilizou-se o servidor *LOMETS* (*LOcal Meta-Threading-Server*) (WU e ZHANG, 2007).

A união dos modelos obtidos a partir de cada técnica foi feita com a ajuda do servidor *AIDA-ab initio domain assembly server* (XU, JAROSZEWSKI e LI, *et al.* 2015). A avaliação dos diedros foi realizada com o programa *PROCHECK*, versão 3.5 (LASKOWSK, MACARTHUR e MOSS *et al.* 1993)

3.2 REFINAMENTO POR DM DO MODELO VWD2-VWD3

O modelo teve seu estado de protonação corrigido para pH 8, predominante no trato intestinal, pelo servidor PDB2PQR (http://nbc222.ucsd.edu/pdb2pqr_2.1.1/). Foi submetido ao estudo de DM para avaliar o comportamento estrutural em função do tempo. Utilizou-se o pacote *GROMACS*, versão 5.1.3 (ABRAHAM, MURTOLA e SCHULZ *et al.* 2015).

As simulações foram realizadas em caixa octaédrica, pois essa geometria permitiu minimizar a quantidade de moléculas de água necessárias para a solvatação. O sistema foi composto de 25288 moléculas de água, o modelo de proteína VWD2-VWD3 e três íons sódio para neutralizar as três cargas negativas do modelo VWD2-VWD3, perfazendo um total de 80525 átomos, sendo 4658 do modelo VWD2-VWD3 e 75864 do solvente.

O sistema foi minimizado utilizando o campo de forças *OPLS-AA* (ROBERTSON e TIRADO-RIVES e JORGENSEN, 2015) implementado no pacote *GROMACS*, versão 5.1.3 (ABRAHAM, MURTOLA e SCHULZ *et al.* 2015). Os algoritmos de minimização de energia utilizados foram, inicialmente, o de máximo declive (*steepest descent*) com restrição de posição com critério de convergência de 600 kJ/mol/nm (14,35 kcal/mol/Å) e, posteriormente, sem restrição de posição com critério de convergência de 300 kJ/mol/nm (7,17 kcal/mol/Å) e, por último, gradientes conjugados com critério de convergência de 150 kJ/mol/nm (3,588 kcal/mol/Å).

Após a minimização da energia do sistema, realizou-se a simulação em duas etapas: primeiro com restrição de posição para átomos pesados, afim de que as águas se ajustassem ao sistema, essa etapa foi de 500 ps a 310 K (temperatura fisiológica de 37° C) com condições periódicas de contorno, e, em seguida, realizou-se a dinâmica propriamente dita, 75 ns também a 310 K sem restrição de posição com passo de integração de 2 fs.

O algoritmo de integração LeapFrog, barostato de Berendsen, termostato V-rescale e algoritmo de restrição de ligação LINCS foram utilizados. Usou-se também condições periódicas de contorno e raios de corte de 1 nm para as interações de

Van der Waals e Coulomb.

O passo a passo para a geração dos arquivos de entrada e execução das minimizações de energias e dinâmicas moleculares e ferramentas de análise das simulações foram realizadas utilizando o pacote *GROMACS*, versão 5.1.3 (ABRAHAM, MURTOLA e SCHULZ *et al.* 2015) e está detalhado no Apêndice A.

3.3 OBTENÇÃO DE DIFERENTES CONFORMAÇÕES DO MODELO VWD2-VWD3 PARA O ESTUDO DE ENSEMBLE *DOCKING*

Utilizando o programa *GMX CLUSTER*, incluído no pacote *GROMACS*, versão 5.1.3, obteve-se um total de quatro conformações do modelo VWD2-VWD3, cada uma dessas representou a média das conformações incluídas dentro de uma faixa de *RMSD* de 2,5 Å. Os passos seguidos para a obtenção das conformações médias estão detalhados no Apêndice A.

3.4 ESTUDO POR *DOCKING* MOLECULAR DAS INTERAÇÕES INTERMOLECULARES DO MODELO VWD2-VWD3 COM QUITOSANAS MODIFICADAS

3.4.1 Desenho, otimização e estado de protonação dos ligantes

Os ligantes estudados foram hexâmeros de cada uma das quitosanas modificadas relacionadas no trabalho de MUELLER, VERROKEN e IQBAL *et al.* (2011), com exceção da N,glutamina-quitosana e apresentadas na Figura 12, sendo eles N,4-tiobutil-amidina-quitosana (TBA-Qui), N,4-ácido mercaptanobenzoico-quitosana (AMB-Qui), N,6-ácido mercaptanonicotínico-quitosana (AMN-Qui), N,acetil-cisteína-quitosana (AC-Qui), N,ácido tioglicólico-quitosana (ATG-Qui) e quitosana (Qui).

Das quatro aminas primárias presentes em cada hexâmero de cada um dos ligantes, uma foi mantida protonada, tornando, assim, a sua carga líquida de +1, exceto o TBA-Qui que exibe nesse caso carga +2, pois apresenta uma amidina protonada. Os modelos foram desenhados utilizando o programa *AVOGADRO*, versão 1.1.1 (HANWELL, CURTIS e CURTIS *et al.* 2012) e subsequentemente suas geometrias

foram otimizadas, clássica e aleatoriamente, com o programa *GHEMICAL* com posterior otimização de geometria pelo programa *MOPAC2016* (método PM7) (STEWART, 2013).

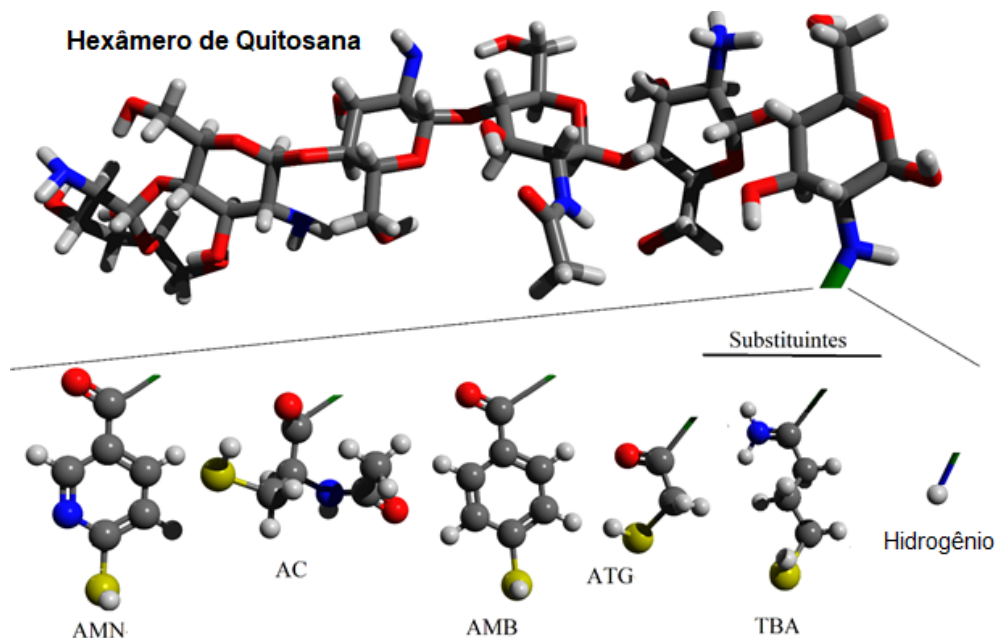


Figura 12. Representação estrutural dos ligantes utilizados no presente trabalho. Imagem gerada a partir do programa *AVOGADRO*, versão 1.1.1 (HANWELL, CURTIS e CURTIS et al. 2012)

3.4.2 Preparação do receptor e ligante para o *docking* molecular

A preparação do ligante foi realizada com o programa *UCSF CHIMERA* versão 1.11.2 (PETTERSEN, GODDARD e HUANG *et al.* 2004), utilizado para a adição de cargas *AM1-BCC*. Após a adição de cargas, o arquivo do ligante foi salvo com extensão *mol2*. Em seguida, esse arquivo foi aberto no programa *AUTODOCKTOOLS* versão 1.4.5 (MORRIS, HUEY e LINDSTROM *et al.* 2009). Para análise dos diedros ativos e, finalmente, salvo com extensão *pdbqt*. O passo seguinte foi abrir o modelo do receptor (*Cluster-1*, *Cluster-2*, *Cluster-3* e *Cluster-4*) selecionar os parâmetros do *grid box*, de modo a envolver toda a macromolécula do receptor. O programa de *docking* utilizado nesta etapa foi o *AUTODOCK VINA* versão 1.1.2 (TROTT e OLSON, 2009).

Foram realizados 10 *dockings* para cada um dos seis ligantes em cada uma das quatro conformações do receptor VWD2-VWD3 utilizando os quatro *clusters*, perfazendo, portanto 240 *dockings*. Dos 40 resultados de *dockings* para cada

ligante, foi selecionado o que apresentou a menor energia livre de interação ($\Delta G_{\text{ligação}}$).

3.4.3 Refinamento dos resultados de *docking* por DM e cálculo de $\Delta G_{\text{ligação}}$ usando a técnica *MMPBSA*

A partir da realização do *docking*, a pose para o complexo receptor-ligante que apresentou o menor $\Delta G_{\text{ligação}}$ foi utilizado como quadro inicial para uma DM do receptor-ligante, seguindo o mesmo protocolo da simulação já realizada na etapa de refinamento do modelo VWD2-VWD3 e descrita no item 3.2. Os ligantes foram parametrizados para o campo de forças OPLS-AA, com os programas ACPYPE versão 0.1.0 (DA SILVA e VRANKEN, 2012) e MKTOP versão 2.2.1 (RIBEIRO, HORTA e ALENCASTRO, 2008). De posse da trajetória da DM do complexo receptor-ligante, utilizou-se o programa *GMXPBSA* versão 2.1.2 (PAISSONI, SPILIOPOULOS e MUSCO *et al.* 2015; PAISSONI, SPILIOPOULOS e MUSCO *et al.* 2014) para acompanhar a evolução do $\Delta G_{\text{ligação}}$ *MMPBSA* ao longo da trajetória.

3.5 RECURSOS COMPUTACIONAIS

Na realização desse trabalho, foram usados os computadores: Laptop Vaio de uso pessoal, processador I5 de terceira geração, 6 GB de memória RAM, 750 GB de HD, placa de vídeo Nvidia GeForce GT 735M; computadores AMD X6 com 16GB de memória RAM e AMD X8, com 32 GB de memória RAM, 1 TB de HD, placa de vídeo PCI-E GeForce GTX970 4GB GDDR5 1664 CUDA cores, adquiridos via aprovação de projeto submetido ao edital 006/2014 – Processo 65656296/2014, localizados no Instituto Federal do Espírito Santo - *Campus Vila Velha*.

4. RESULTADO E DISCUSSÃO

4.1 OBTENÇÃO DO MODELO MUC2

Dentro da sequência de 314 aminoácidos, o servidor *RaptorX* modelou dois domínios, o primeiro com resíduos de número 107 a 208 e o segundo com resíduos 209 a 314, ambos tiveram como *template* a estrutura terciária sob identificação

2mhp:A. Essa identificação no *PDB* representa o *ensemble* estrutural obtida por técnica de RMN do domínio D do FvW, com identidades de sequência de 27% e 32%, respectivamente. Além da identidade de sequência, o servidor fornece mais alguns parâmetros de avaliação do modelo, entre eles, o *SCORE* que é o resultado do alinhamento da sequência primária e pode ir de 0 até um valor máximo igual ao número de resíduos da sequência primária a ser modelada, sendo que, quanto mais próximo de 0 menor é a qualidade do modelo. O parâmetro *GDT* (*Global Distance Test*) ou teste de distância global, o *GDT* é um algoritmo de comparação de estruturas 3-D que usa iterações sucessivas dos resultados de sobreposição de resíduos presentes nas estruturas do *template* e dos modelos a serem avaliados. Essas sobreposições seguem os seguintes passos: primeiro sobrepõem-se as estruturas, computando somente aqueles resíduos que caem dentro de um raio de corte de *RMSD* inicialmente baixo de 1 Å, por exemplo, excluindo todos os outros, em seguida, aumenta-se o raio de corte, incluindo resíduos que ficaram de fora na etapa anterior, esses procedimentos vão sendo executados até que nenhum resíduo seja mais excluído. O *GDT* é obtido pela Equação 16:

$$GDT = \frac{(C_1 + 0,75C_2 + 0,5C_3 + 0,25C_4)}{N} \cdot 100 \quad (16)$$

Em que C_1 , C_2 , C_3 e C_4 representam os números de resíduos que caem dentro dos raios de corte 1, 2, 4 e 8 Å, respectivamente, e N o número de resíduos da cadeia proteica. Para modelos com número de resíduos menores que 100 aminoácidos um *GDT* acima de 50 é um bom indicador (KÄLLBERG, WANG, H. e WANG, S. *et al.* 2012, LANCIA e ISTRAIL, 2003).

Um último parâmetro de avaliação e classificação utilizado pelo servidor é o *P-value*, esse parâmetro representa a probabilidade de o modelo predito ser pior que o melhor modelo de um conjunto gerado aleatoriamente. Quanto menor um valor *P-value* melhor será o modelo, *P-value* menores que 10^{-3} e 10^{-4} para α -hélices e para folhas- β , respectivamente, significam bons modelos. A tabela 2 traz os valores de *SCORE*, *GDT* e *P-value* para os dois domínios modelados pelo servidor *RAPTORX*.

Tabela 2. Parâmetros de classificação para os modelos do *RaptorX*

MODELO (resíduos)	107-208	209-314
SCORE	69	72
GDT	62	65
P-value	1.60×10^{-9}	8.81×10^{-13}

Pelos valores dos itens de avaliação apresentados na Tabela 2, observa-se que os dois domínios gerados pelo servidor RaptorX satisfazem os critérios para serem classificados como bons modelos a serem utilizados posteriormente.

Como o domínio gerado pelo RaptorX para os 106 primeiros resíduos não tiveram uma boa identidade de sequência com o *template* selecionado, recorreu-se a uma técnica de modelagem mais apropriada para casos de baixa identidade de sequência que se trata da técnica de *Threading*.

LOMETS é um servidor que emprega algoritmos de nove outros servidores integrados em um só local, dedicados ao alinhamento e seleção de *templates* para modelos 3D de proteínas pelo método de *Threading*. Dentre os melhores modelos previstos por cada um dos nove servidores são selecionados aqueles que apresentem consenso. Nessa etapa, os *templates* são compostos apenas de *C α* .

Após a seleção dos melhores *templates*, utiliza-se o programa *MODELLER* para a inserção das cadeias laterais (WU e ZHANG, 2007). O primeiro modelo na classificação do servidor *LOMETS*, bem como os modelos do servidor *RaptorX*, estão representados na Figura 13.

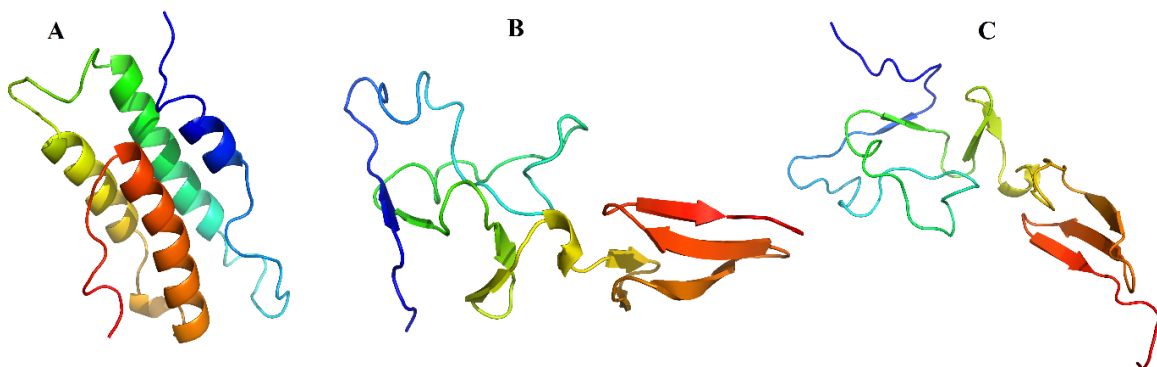


Figura 13. Domínios modelados, em A 106 primeiros resíduos modelados pelo servidor *LOMETS* por *Threading*, em B e C resíduos de 107 a 208 e 209 a 314, respectivamente, por modelagem comparativa usando o *RaptorX*. Imagem gerada pelo programa *PYMOL* versão 1.2r3pre (Schrödinger, LLC)

Os três domínios criados na etapa anterior precisaram ser unidos na sequência correta, dando origem ao modelo final, para essa finalidade usou-se um servidor que cumpriu esse objetivo. O servidor *AIDA* (XU, JAROSZEWSKI e LI, *et al.* 2015) promove a conexão entre os modelos sem alterar a conformação interna de cada domínio. Dessa forma, a posição e consequentemente a energia associada aos resíduos dentro dos domínios permanecem constantes, sofrendo variação somente os resíduos adjacentes aos sítios de ligação entre os domínios. O servidor ainda disponibiliza um executável de fácil operação que renumera os aminoácidos dos arquivos de extensão *pdb* dos domínios de acordo com a sequência completa de aminoácidos contidos no arquivo de extensão *fasta*, pois é necessário que os domínios estejam enumerados em sequência para que sejam ligados corretamente. Na Figura 14, apresenta-se o modelo completo referente aos 314 aminoácidos da região entre os domínios VWD2-VWD3.

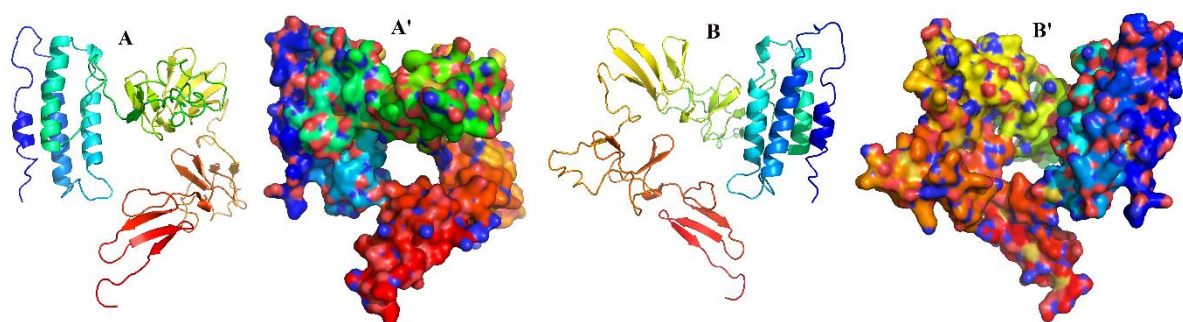


Figura 14. Modelo completo VWD2-VWD3 em diferentes perspectivas, obtido pela união dos domínios, (FIGURA 13), pelo servidor *AIDA*, A e B representação em fita, A' e B' em superfície. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

Como esperado em função da escolha do segmento de aminoácidos, o modelo apresentou 10 cisteínas e 7 cistinas. Como visto na Figura 15.

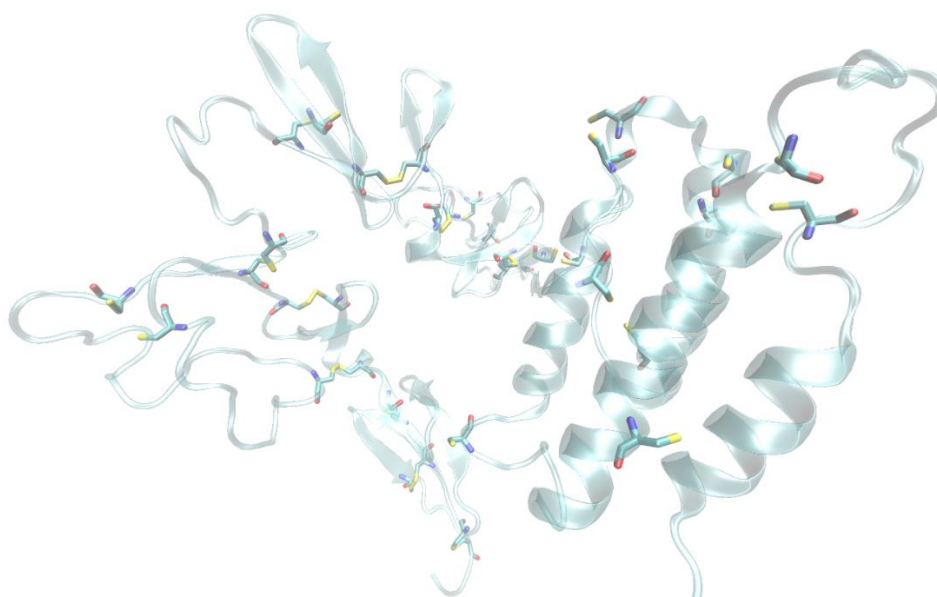


Figura 15. Representação em fita do modelo com ênfase nas cistinas e cisteínas

Além dos parâmetros de avaliação do servidor, existem outros que dão uma visão mais precisa da qualidade dos modelos em diferentes aspectos. Para avaliar a qualidade estereoquímica de um modelo é amplamente utilizado o Diagrama de Ramachandran, ele consiste na disposição, em um plano, dos valores assumidos pelos ângulos diedros ψ versus ϕ . Os ângulos ψ representam as rotações em torno da ligação $C\alpha-C$ e os ângulos ϕ em torno da ligação $N-C\alpha$ de um determinado resíduo na cadeia peptídica, conforme mostrado na Figura 16.

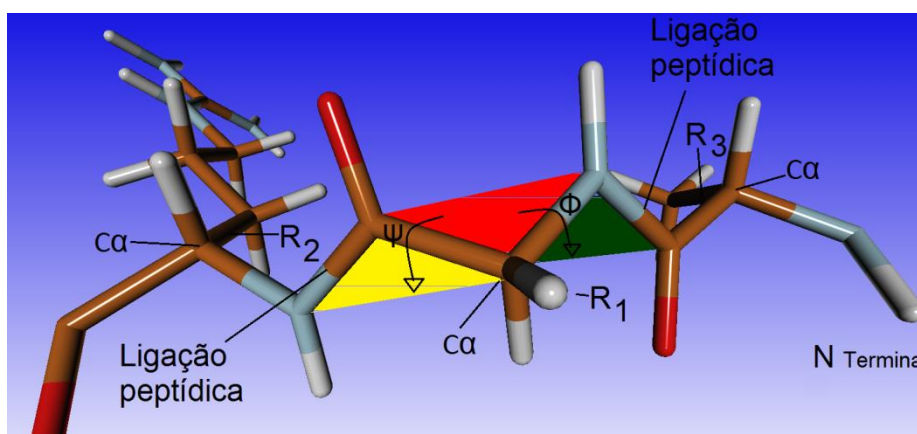
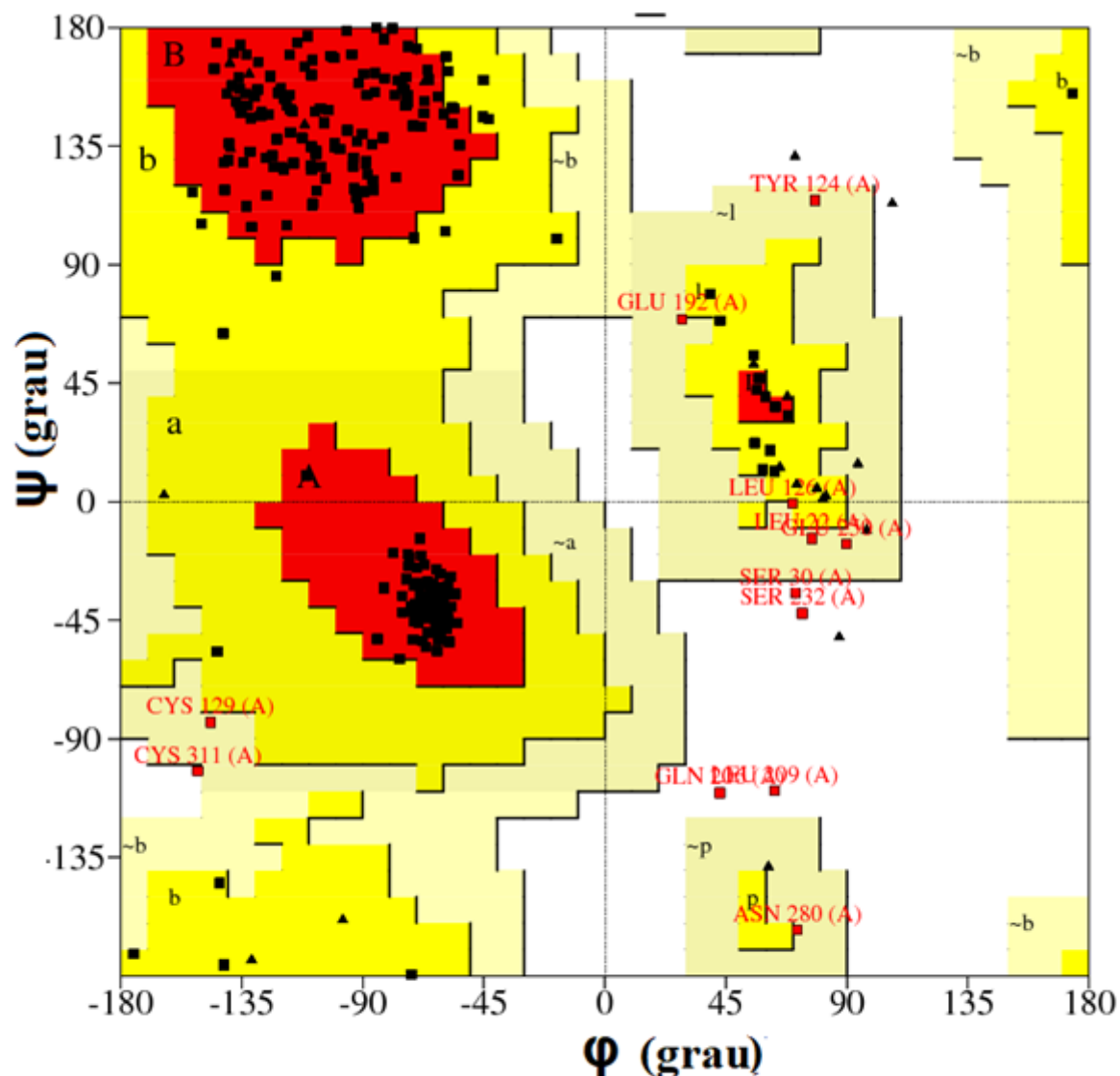


Figura 16. Ilustração dos ângulos diedros ψ e ϕ em uma cadeia tripeptídica, R1 (glicina), R2 (arginina) e R3 (alanina). Imagem gerada pelo programa *GMOLDEN*, versão 5.0 (SCHAFTENAAR e NOORDIK, 2000) e adaptado pelo autor

O diagrama foi nomeado em homenagem ao pesquisador que evidenciou haver um padrão no modo como os ângulos diedros ψ e ϕ se comportam quando plotados em um gráfico de ϕ versus ψ (FIGURA 17).



Resíduos em regiões mais favoráveis [A,B,L]	243	87,4%
Resíduos em regiões adicionalmente permitidas [a,b,l,p]	23	8,3%
Resíduos em regiões generosamente permitidas [~a,~b,~l,~p]	7	2,5%
Resíduos em regiões proibidas	5	1,8%

Figura 17. Diagrama de Ramachandran, gerado pelo *PROCHECK*, versão 3.5 (LASKOWSK, MACARTHUR e MOSS et al. 1993), do modelo VWD2-VWD3 obtido por modelagem comparativa e *Threading*

Isso mostrou que os aminoácidos em uma proteína tendem a ocupar regiões bem definidas quanto à combinação desses ângulos. Quando se obtém a combinação de ângulos diedros da cadeia principal de inúmeras estruturas já depositadas em banco de dados, observa-se que tendem a se concentrarem nas regiões muito favoráveis

do diagrama de Ramachandran sinalizadas com **A**, **B** e **L**, que são justamente os ângulos assumidos pelos aminoácidos quando estão compondo os diferentes tipos de estruturas secundárias α -hélices, folhas- β e α -hélices, respectivamente, regiões vermelhas da Figura 17, quanto maior a resolução da técnica de determinação da estrutura terciária, maior é a densidade de resíduos que caem nas regiões muito favoráveis (**A**, **B** e **L**) (SCHLICK, 2006). Isso demonstra um padrão seguido pelas proteínas em sua estrutura que tem como resultado a diminuição da energia potencial, decorrentes da minimização do impedimento estérico dos átomos como esferas de Van der Waals. De acordo com o programa *PROCHECK* versão 3.5 (LASKOWSK, MACARTHUR e MOSS *et al.* 1993), o modelo VWD2-VWD3 apresentou 87,4% dos resíduos nessas regiões.

Outro critério de avaliação usado foi o *Verify3D* (LÜTHY, BOWIE e EISENBERG, 1992), um servidor (http://services.mbi.ucla.edu/Verify_3D/) que usa abordagem estatística a partir de banco de dados contendo 18 ambientes com perfis de estruturas secundárias expostas ao solvente compiladas a partir de estruturas de alta resolução. A qualidade da proteína é, então, avaliada para cada aminoácido, verificando a compatibilidade de cada resíduo do modelo com as pontuações obtidas para cada resíduo das estruturas de referências. Pontuação de resíduo abaixo de zero é considerada um indicador de baixa qualidade, dessa forma é possível determinar qual parte do modelo contribui para sua perda de qualidade (XIONG, 2006: 220). A saída é um gráfico em duas dimensões relacionando resíduo versus pontuação como no resultado para o modelo VWD2-VWD3 (FIGURA 18).

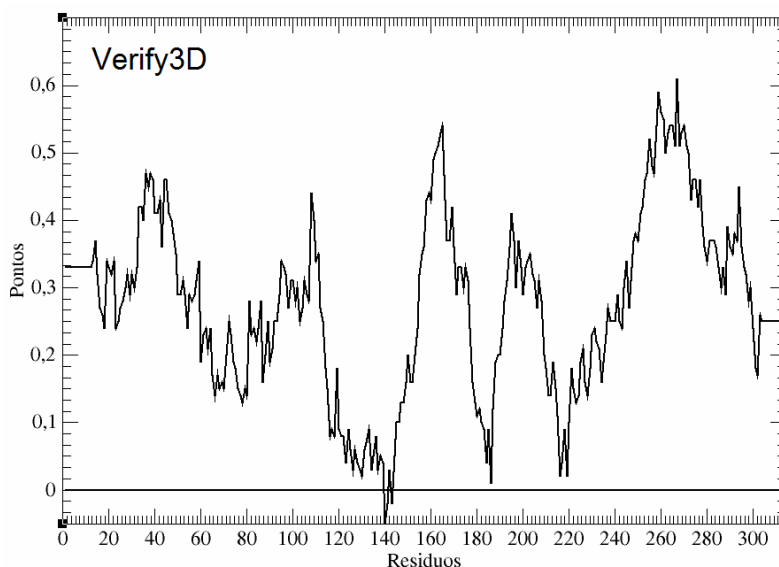


Figura 18. Resultado da submissão do modelo VWD2-VWD3 ao servidor *Verify3D*

Como se pode observar, o modelo apresenta uma boa avaliação apresentando restrições nos resíduos 140 (ácido aspártico), 141 (serina) e 143 (cisteína). O modelo obtido nessa etapa servirá de ponto de partida para os demais passos da pesquisa, submetido inicialmente à simulação por DM.

4.2 ANÁLISE DA DM NA ETAPA DE REFINAMENTO DO MODELO VWD2-VWD3

A primeira análise básica feita foi a partir do *RMSD*, esse parâmetro é obtido por meio da sobreposição das conformações, dadas pela trajetória, a uma estrutura de referência, ele é uma medida do quanto as conformações mudam ao longo de uma simulação. Sendo descrito pela Equação 17.

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (17)$$

Onde $M = \sum_{i=1}^N m_i$ e $r_i(t_1)$ é a posição de cada átomo i no tempo t .

A Figura 19 traz a variação do *RMSD* ao longo do tempo de simulação para os átomos que compõem o esqueleto peptídico do modelo VWD2-VWD3, a saber, os 942 átomos compostos por N e C das ligações peptídicas bem como o $C\alpha$, a

referência usada foi aquela do tempo $t = 0$, ou seja, o modelo VWD2-VWD3 tal como obtido na etapa de modelagem estrutural.

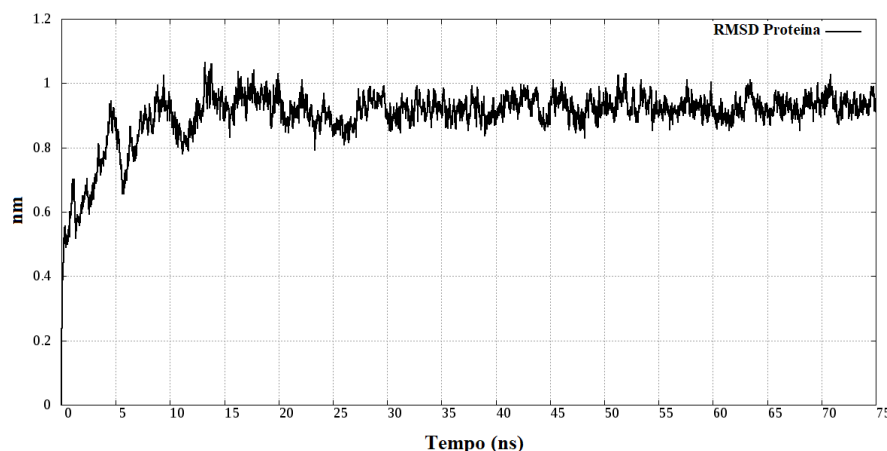


Figura 19. Gráfico *RMSD* versus Tempo de simulação do modelo VWD2-VWD3

Quando a estrutura de referência é a estrutura inicial, ou seja, a conformação assumida no tempo $t = 0$, o *RMSD* tende a aumentar no transcorrer da simulação, pois há uma mudança de conformação após o início da dinâmica, porém, a partir de um determinado tempo de simulação, o *RMSD* converge para um certo valor, no caso da Figura 19, em torno de 1 nm. A observação dessa convergência é um indicador de que o modelo está se estabilizando e entrando em uma condição de equilíbrio, condição fundamental para que se possam obter propriedades do sistema com base na mecânica estatística. Observa-se que o modelo VWD2-VWD3 tem seu *RMSD* se estabilizando já a partir dos 30 ns. Um ponto a se notar é que o modelo, após o início da simulação, sofre uma variação acentuada do seu *RMSD*, chegando a aproximadamente 10 Å, o que mostra um grande distanciamento da conformação da estrutura inicial.

Na Figura 20, mostra-se a sobreposição da conformação inicial (verde) e a conformação assumida após os 75 ns (azul) de simulação.

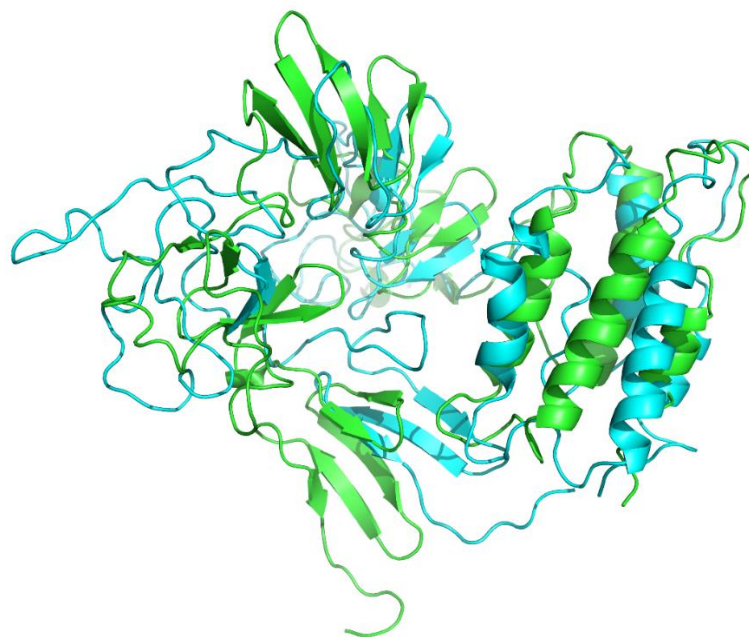


Figura 20. Sobreposição da conformação inicial (verde, $t = 0$ ns) e conformação final (azul, $t = 75$ ns) do modelo VWD2-VWD3, representação em fita. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

Outro parâmetro que permite avaliar a estabilidade da proteína é o Fator-B ou Fator de Temperatura, esse conceito provém da cristalografia de raios-X e mostra a incerteza associada à posição de um átomo no cristal, dessa forma, quanto maior a variação da posição, maior será o Fator-B (VERLI, 2014: 280). Na Figura 21, observamos a representação do Fator-B em duas formas diferentes: em azul região na qual há menor flutuação em torno da estrutura inicial no $t = 0$ da DM, já em vermelho, maior flutuação.

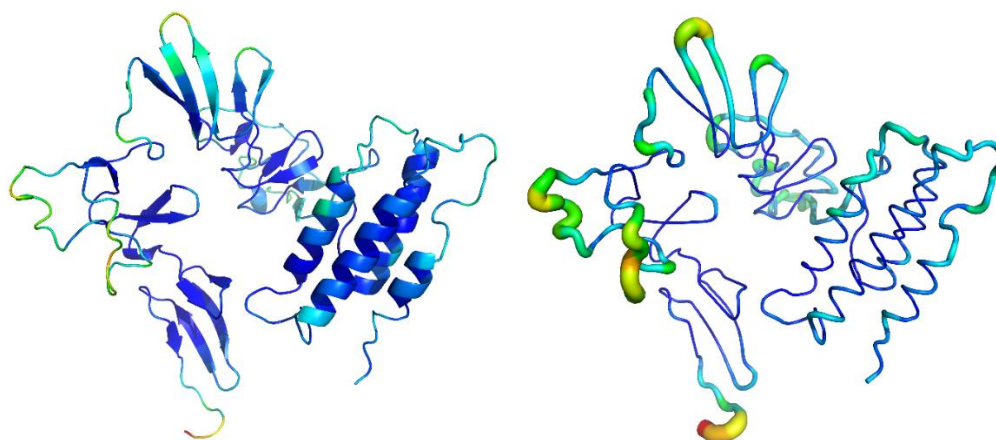


Figura 21. Representação para o Fator-B de cada aminoácido do modelo VWD2-VWD3, representação (esquerda) em fita e massas (direita). Do azul para o vermelho em ordem crescente de oscilação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

A DM também permite acompanhar as interações do tipo ligação hidrogênio ao longo do tempo (FIGURA 22), os critérios utilizados para o mapeamento dessas interações foram de 3,5 Å para o raio de corte e 60° para o ângulo de corte. As interações por ligação hidrogênio entre os resíduos de aminoácidos são um dos principais fatores responsáveis pela estabilização da estrutura proteica, em termos gerais, pode-se considerar que para cada ligação intramolecular estabelecida, outra com o solvente é desfeita (VERLI, 2014: 151).

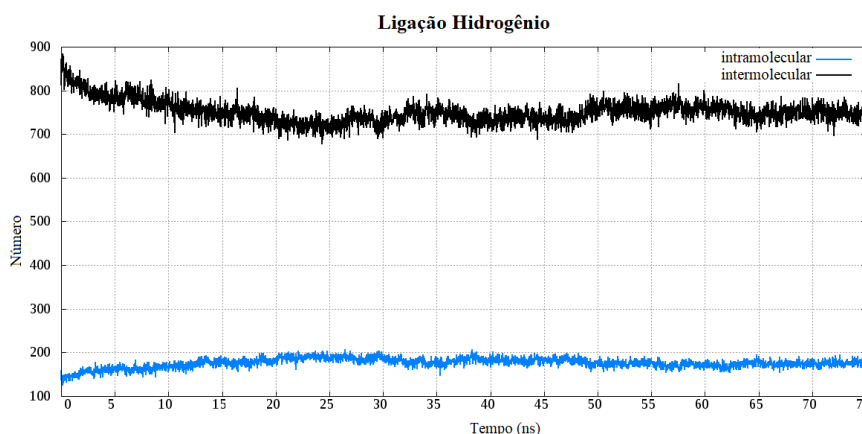


Figura 22. Evolução das ligações hidrogênio inter e intracadeia ao longo da simulação por DM

Como se pode notar na Figura 22, houve, ao longo da simulação, um incremento nesse tipo de interação intramolecular.

Além das ligações hidrogênio, outra interação não ligada responsável pela estabilização da estrutura nativa das proteínas são os contatos hidrofóbicos. Esses contatos são impelidos pela diminuição de organização das moléculas de água em torno dos resíduos hidrofóbicos. Por sua vez, a ordenação dos aminoácidos apolares no interior da estrutura diminui o número de aminoácidos expostos ao solvente e, portanto, a Área Superficial Acessível ao Solvente (ASAS), ou seja, a camada de solvatação. O efeito desse fenômeno é o aumento da entropia do solvente (VERLI, 2014: 151). As ferramentas de análise da DM permitem ter estimativa da ASAS. O que o algoritmo faz é, basicamente, varrer a superfície da proteína com um elemento de prova consistindo em uma esfera cujo raio é igual ao raio de Van der Waals para a molécula do solvente, então, a área formada pelos pontos coincidentes com o centro da esfera de Van der Waals do elemento de prova é a ASAS (FIGURA 23).

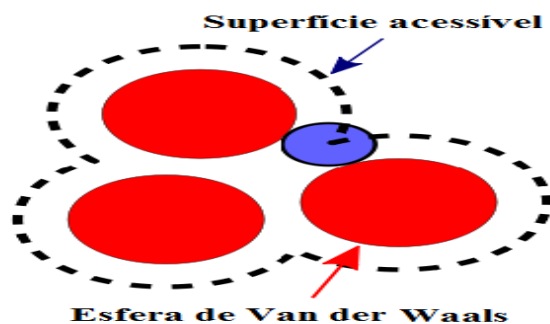


Figura 23. Ilustração da ASAS. Adaptado da *Wikipédia*

A ASAS para a modelo de MUC2 exibida na Figura 24 mostra uma diminuição acentuada até 17 ns seguida por uma estabilização entre 180 a 200 nm². Esse comportamento evidencia maior compactação da estrutura no transcorrer da DM, decorrente das interações de natureza hidrofóbicas.

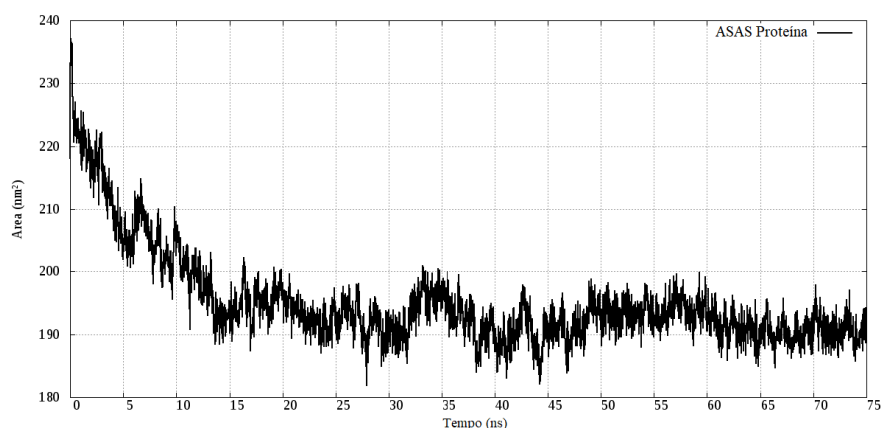


Figura 24. ASAS ao longo da simulação por DM

O próximo passo na condução da pesquisa foi o estudo das interações químicas entre o receptor VWD2-VWD3 e os modelos de quitosanas modificadas, isso se deu por meio da técnica de *docking* molecular.

Existem algumas estratégias para contornar a limitação da rigidez do receptor na técnica de *docking*, a que foi usada aqui é o conceito de *cluster*. *Cluster* é uma conformação que representa a média de um conjunto de conformações assumidas durante a simulação de DM. Na Figura 25, apresentam-se os quatro *clusters* obtidos para o estudo de *docking*.

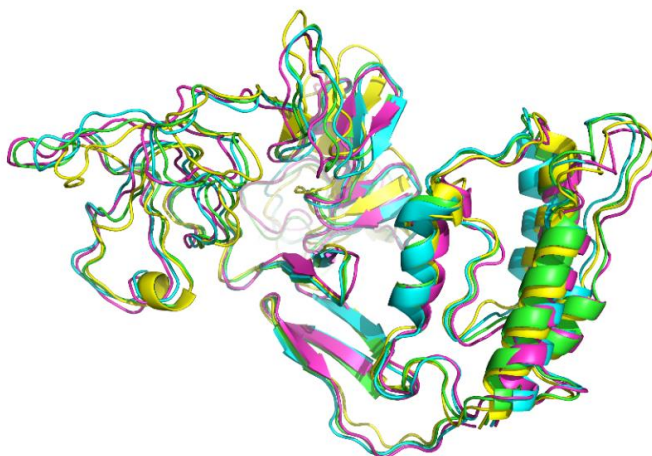


Figura 25. Sobreposição dos 4 *clusters* gerados a partir dos últimos 10 ns de simulação com um raio de corte de 2.5 Å. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

O *Cluster-1* (verde) representou a conformação média de 449 conformações cujos *RMSD* variaram dentro 2,5 Å, o *Cluster-2* (azul) 36 conformações, *Cluster-3* (roxo) 9 conformações e, finalmente, o *Cluster-4* (amarelo) 7 conformações.

A escolha das conformações possíveis dentro da faixa dos últimos 10 ns de simulação se deram em função da grande quantidade de clusters que eram gerados quando se ampliava a faixa de simulação. Inviabilizando a pesquisa, diante do custo computacional que isso demandaria.

4.3 RESULTADOS DE *DOCKING* PARA OS HEXÂMEROS DE QUITOSANA E DIFERENTES QUITOSANAS MODIFICADAS

Os menores valores para $\Delta G_{(ligação-docking)}$ entre os ligantes e as diferentes conformações para o receptor VWD2-VWD3 estão dispostos na Tabela 3:

Tabela 3. Resultados de *docking* para cada ligante com seu respectivo *cluster*

LIGANTE	Cluster 1	Cluster 2	Cluster 3	Cluster 4	$\Delta G_{(ligação-docking)}$ (kJ/mol)
AC-Qui	X				-33,4
AMB-Qui			X		-35,9
AMN-Qui				X	-35,5
ATG-Qui		X			-34,7
Qui			X		-34,3
TBA-Qui		X			-33,9

Pelos valores de $\Delta G_{(ligação-docking)}$ dos ligantes exibidos na Tabela 3, observa-se não ser possível diferenciá-los entre si, quanto a crescente estabilidade do complexo, pois o erro padrão de 11,7 kJ/mol associado ao programa *AUTODOCK VINA* versão 1.1.2 (TROTT e OLSON, 2009) mostra que não é possível distingui-los, estatisticamente, entre os valores para $\Delta G_{(ligação-docking)}$ dos complexos, de modo que, fez-se necessário o uso da metodologia *MMPBSA*.

As Figuras 26 a 31 mostram os melhores quadros ou pose para proteína e ligantes AC-Qui, AMB-Qui, AMN-Qui, ATG-Qui, Qui e TBA-Qui respectivamente. As descrições mais detalhadas das interações receptor-ligantes constam no Apêndice B.

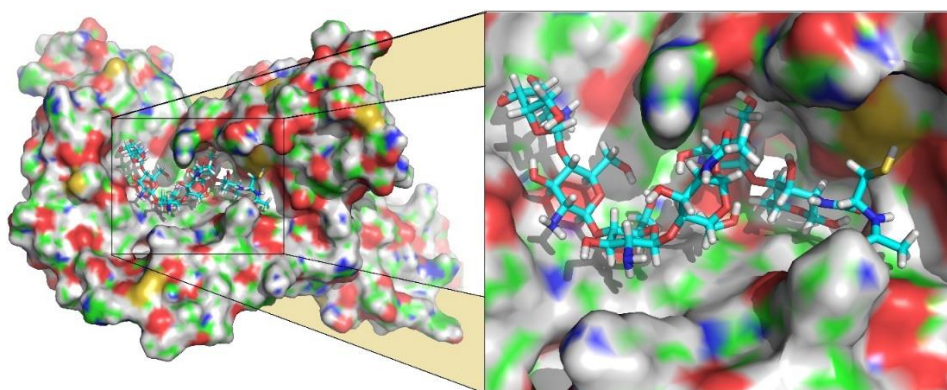


Figura 26. Melhor pose para o ligante AC-Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

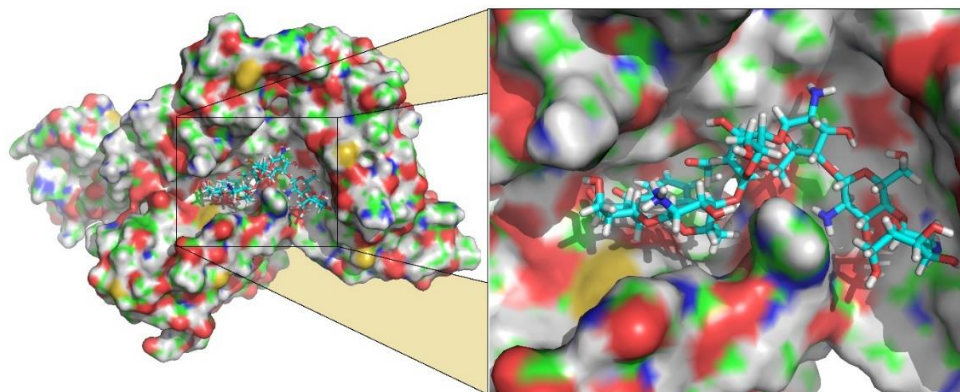


Figura 27. Melhor pose para o ligante AMB-Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

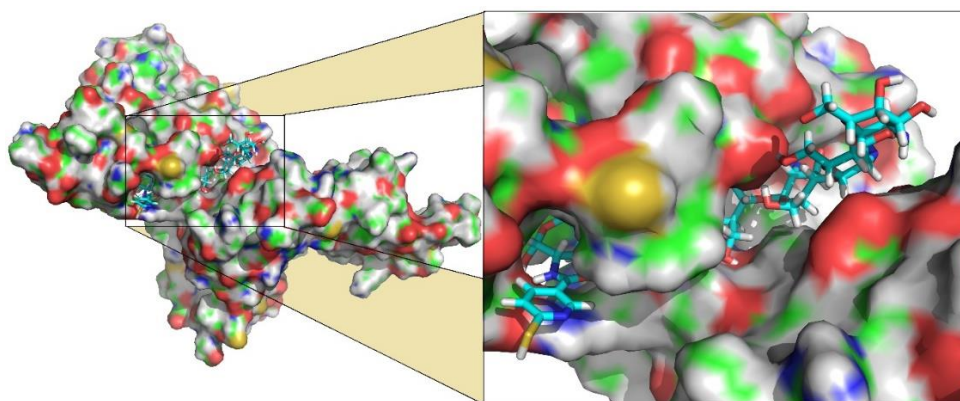


Figura 28. Melhor pose para o ligante AMN-Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

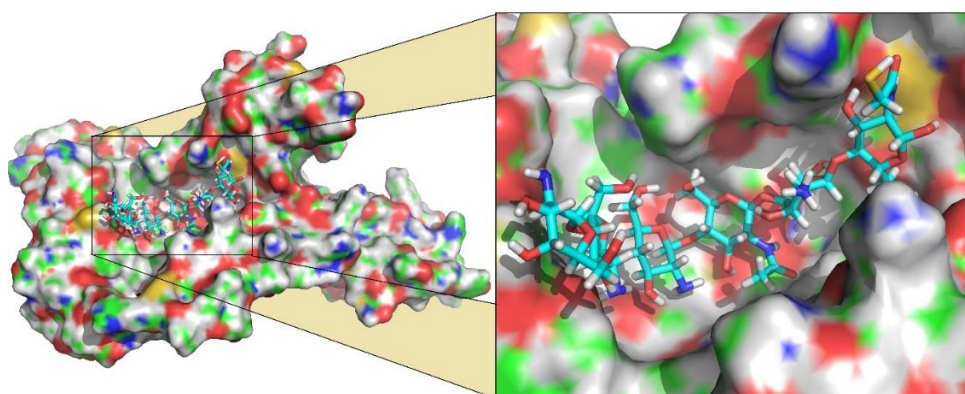


Figura 29. Melhor pose para o ligante ATG-Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

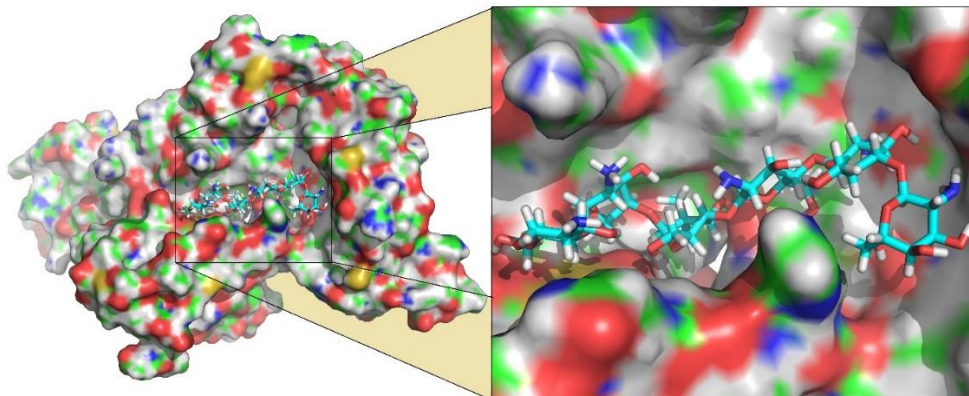


Figura 30. Melhor pose para o ligante Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

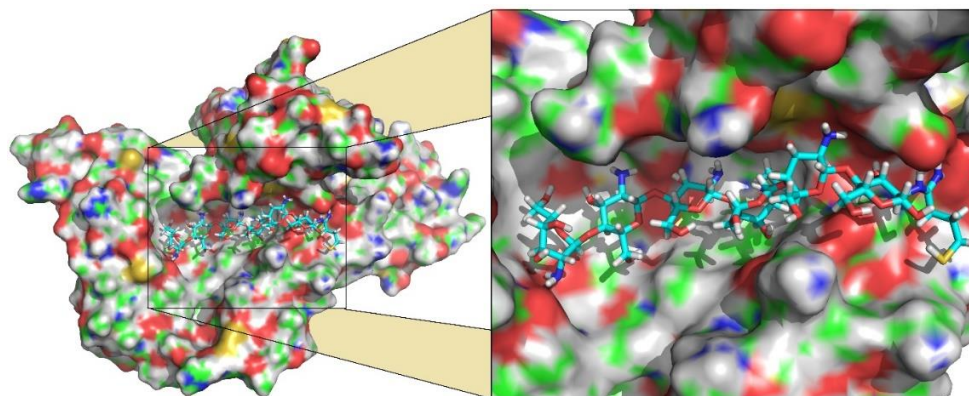


Figura 31. Melhor pose para o ligante TBA-Qui, obtido a partir do *docking* molecular. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

A importância da técnica de *docking* nesta pesquisa se deu à medida que ela permitiu obter as melhores configurações iniciais, para a interação receptor-ligante, utilizada na realização de DM posterior, a partir da qual foi possível obter uma série de elementos de análise, além do $\Delta G_{(ligação-MM/PBSA)}$ pelo método *MMPBSA*.

4.4 ANÁLISE DA DM DO COMPLEXO RECEPTOR-LIGANTE E $\Delta G_{ligação}$ OBTIDO PELO MÉTODO *MMPBSA*

As simulações de 15 ns para cada sistema contendo receptor e ligante, a partir do quadro de *docking* de menor energia para o complexo, geraram as trajetórias usadas nas análises feitas nessa seção, bem como a obtenção do $\Delta G_{(ligação-MM/PBSA)}$.

Nas Figuras 32 a 37, observa-se a evolução da conformação do ligante e do sítio de interação ao longo dos 15 ns.

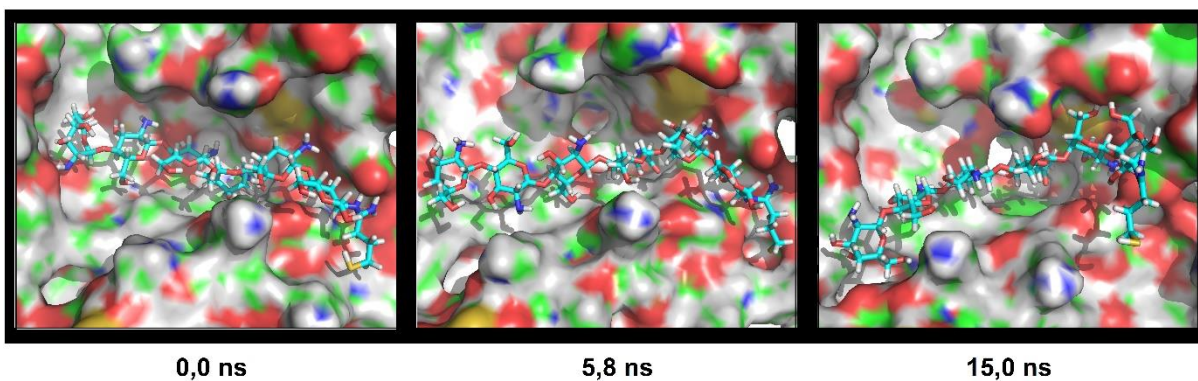


Figura 32. Mudança do ligante TBA-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

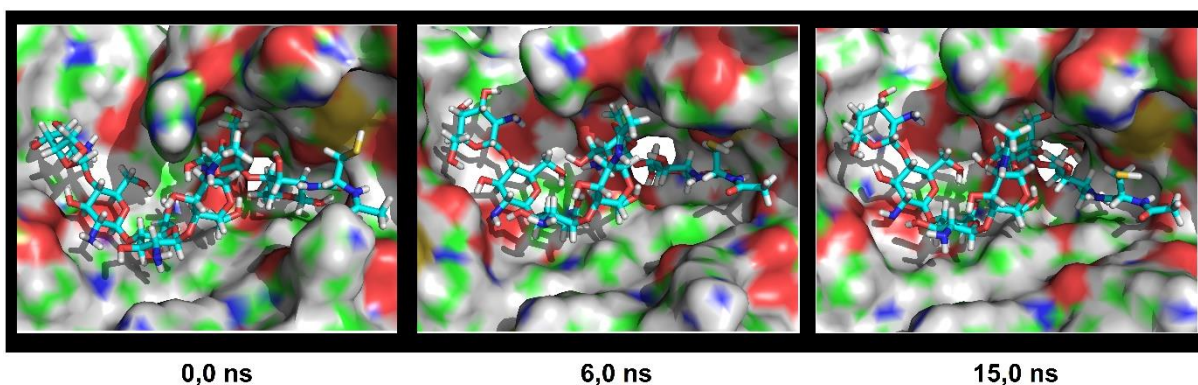


Figura 33. Mudança do ligante AC-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

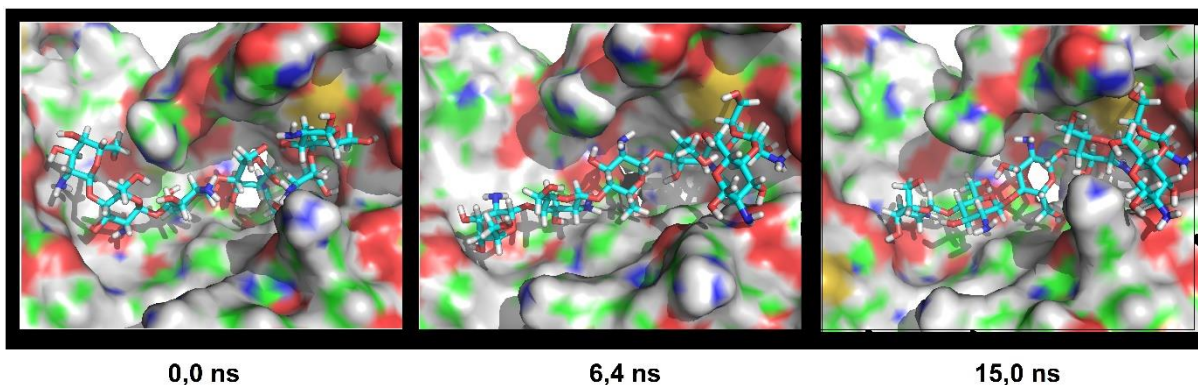


Figura 34. Mudança do ligante Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

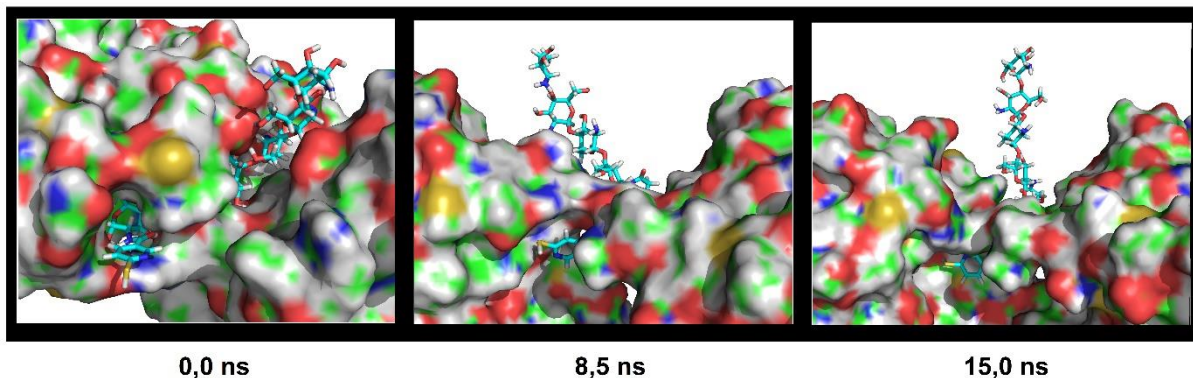


Figura 35. Mudança do ligante AMN-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

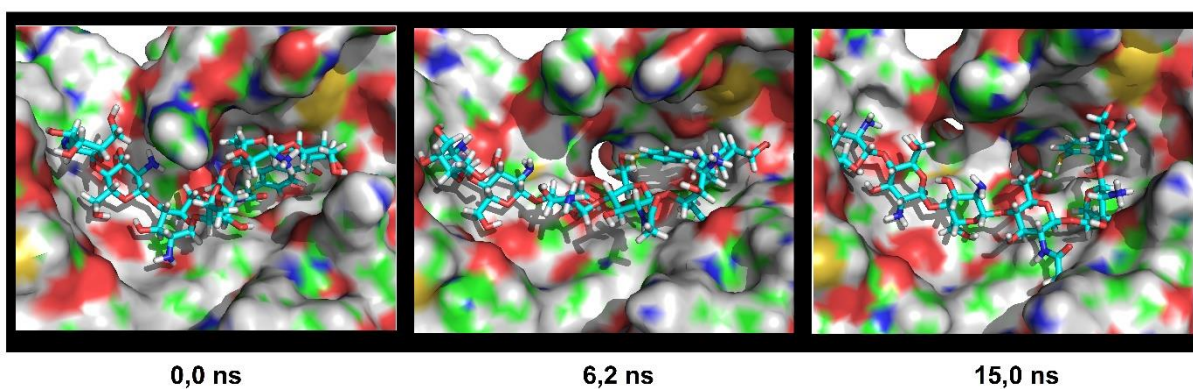


Figura 36. Mudança do ligante AMB-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

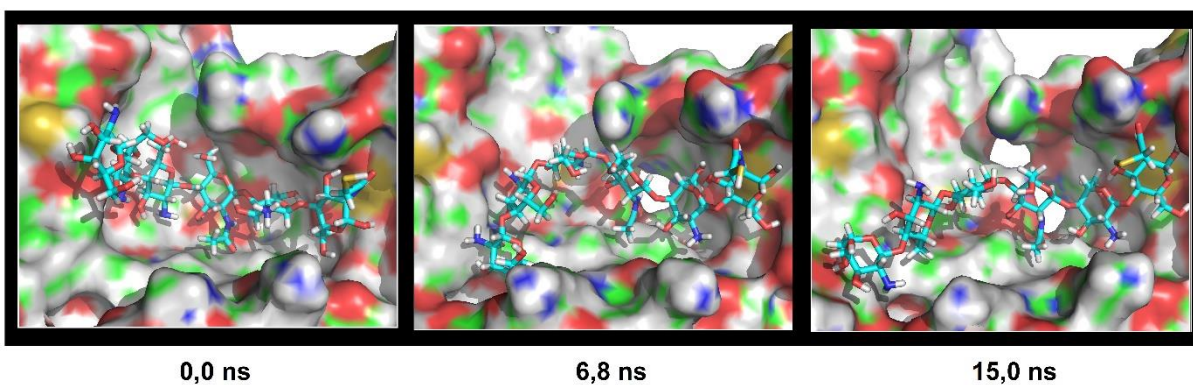


Figura 37. Mudança do ligante ATG-Qui ao longo dos 15 ns de simulação. Imagem gerada pelo programa *PYMOL*, versão 1.2r3pre (Schrödinger, LLC)

Os valores do $\Delta G_{(ligação-MM/PBSA)}$ ao longo do tempo para as simulações por DM estão dispostos na Figura 38.

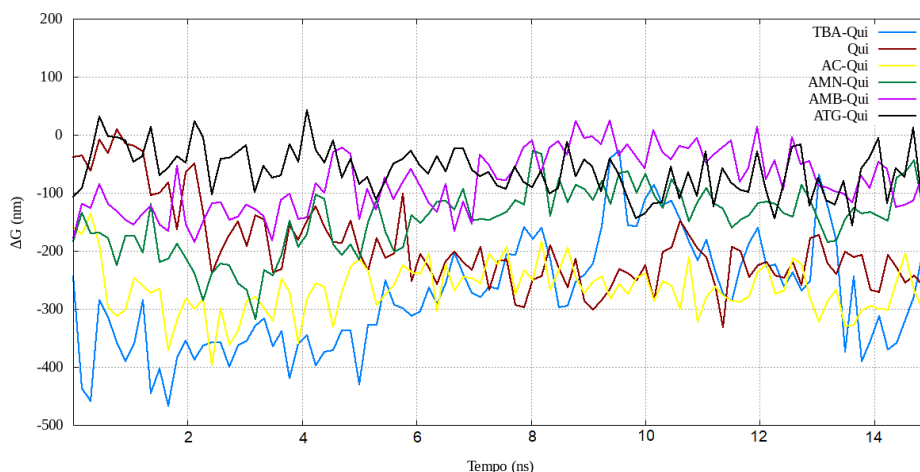


Figura 38. $\Delta G_{(ligação-MM/PBSA)}$ receptor-ligante ao longo de 15 ns de simulação

Determinar o ΔG de qualquer processo ao longo do tempo só é possível por meio da técnica de DM, uma vez que as medidas das técnicas experimentais nos informam sobre médias no tempo.

Quando se consideram as médias no tempo (TABELA 4), os ligantes TBA-Qui e AC-Qui figuram com os melhores valores de $\Delta G_{(ligação-MM/PBSA)}$, sendo ambos estatisticamente idênticos. O ligante Qui apresentou o terceiro menor valor de $\Delta G_{(ligação-MM/PBSA)}$, seguidos por AMN-Qui, AMB-Qui e ATG-Qui.

Uma vantagem da técnica *MMPBSA* é a possibilidade de se explicitar a contribuição de cada um dos principais tipos de interações.

A decomposição do $\Delta G_{(ligação-MM/PBSA)}$ com cada termo contribuinte estão dispostos na Tabela 4.

Tabela 4. Média dos resultados de $\Delta G_{(ligação-MM/PBSA)}$ para cada ligante ao longo da simulação por DM, com detalhes da contribuição de cada termo para o resultado final. Ordem decrescente da esquerda para a direita

TERMO (kJ/mol)	TBA-Qui	AC-Qui	QUI	AMN-Qui	AMB-Qui	ATG-Qui
ΔE_{el}	-1321,877	-703,148	-807,185	-570,708	-211,766	-361,951
ΔE_{vdW}	-205,174	-306,593	-202,070	-195,588	-257,828	-197,313
ΔG_{polar}	1284,535	776,332	826,647	647,684	421,205	520,442
ΔG_{apolar}	-30,485	-36,182	-28,597	-24,973	-32,380	-29,463
$\Delta G_{(ligação)}$	-273,001	-269,592	-211,204	-143,585	-80,700	-68,285
Erro padrão	3,559	1,727	1,985	2,004	2,004	1,715

Em uma consideração geral sobre os dados da Tabela 4, os valores negativos são responsáveis pela estabilização do complexo, já valores positivos desestabilizam-no, pode-se observar, portanto, que a influência de natureza polar do solvente é um grande fator desestabilizador. A água, um solvente polar, substitui ligações hidrogênio receptor-ligante por receptor-água e ligante-água. Além de blindar as interações eletrostáticas receptor-ligante, em razão da sua alta constante dielétrica (NELSON e COX, 2014).

Para o complexo TBA-Qui-VWD2-VWD3, observa-se que as interações de natureza eletrostática são as grandes responsáveis pela estabilização do complexo, lembrando que o grupo amidina protonada do ligante TBA-Qui aumenta a carga líquida positiva da quitosana modificada, favorecendo esse tipo de interação com o receptor, um fato já apontado por MUELLER, VERROKEN e IQBAL *et al.* (2011) para justificar o melhor desempenho de mucoadesão para o TBA-Qui em face dos outros ligantes. Além de exibir o maior número de ligação hidrogênio, 9.886, em média, contra 9.777 AC-Qui, 7.344 Qui, 5.296 AMN-Qui, 4.067 AMB-Qui e 6.096 para o ATG-Qui, (FIGURA 39).

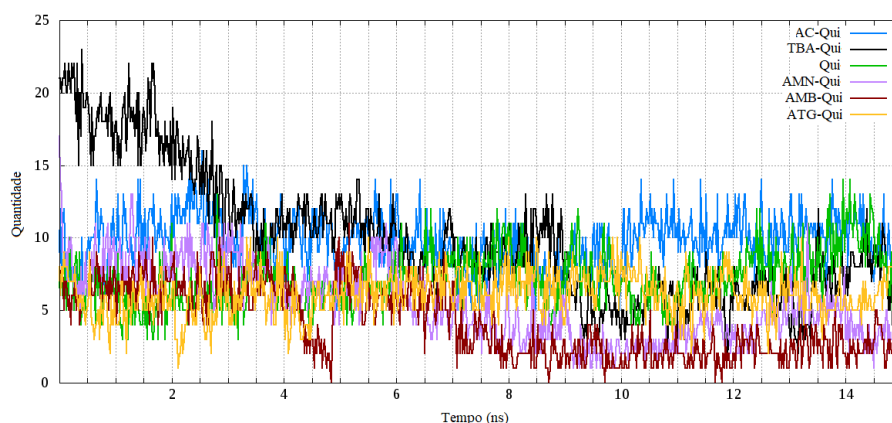


Figura 39. Número de ligação hidrogênio receptor-ligante em função do tempo

O ligante TBA-Qui também apresentou prolongadas interações iônicas entre a amidina protonada e os resíduos glutamato-274 e aspartato-157, que pode ser observado na Figura 40

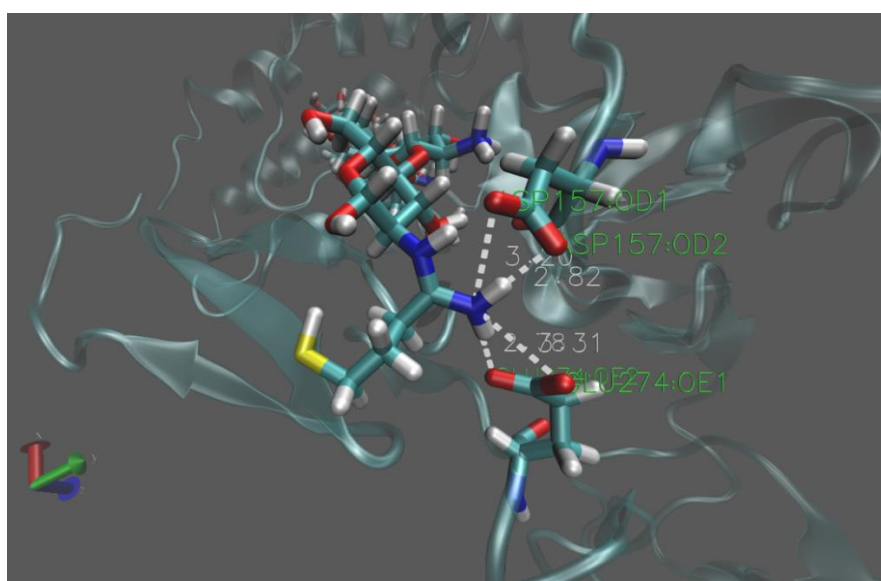


Figura 40. Pontes salinas entre amidina protonada do ligante TBA-Qui, aspartato (ASP) 157 e glutamato (GLU) 274. Imagem gerada pelo programa VMD, versão 1.9.3 (HUMPHREY, DALKE e SCHULTEN, 2012)

As distâncias entre a amidina protonada do ligante TBA-Qui e os resíduos aspartato (ASP) 157 e glutamato (GLU) 274 ao longo da simulação estão descritas na Figura 41.

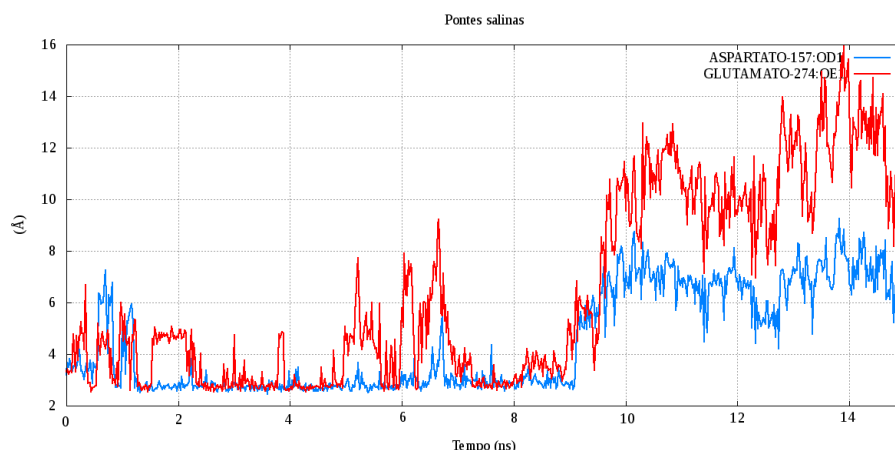


Figura 41. Distâncias entre amidina protonada do ligante TBA-Qui, aspartato (ASP) 157 e glutamato (GLU) 274

Quando se avaliou somente as contribuições do substituinte no hexâmero para as ligações hidrogênio receptor-ligante (FIGURA 42), observou-se que o TBA-Qui, estabeleceu maior número desse tipo de interação 1.480, em média, contra 1.207 AC-Qui, 0.000 Qui, 0.174 AMN-Qui, 0.000 AMB-Qui e 0.713 para o ATG-Qui.

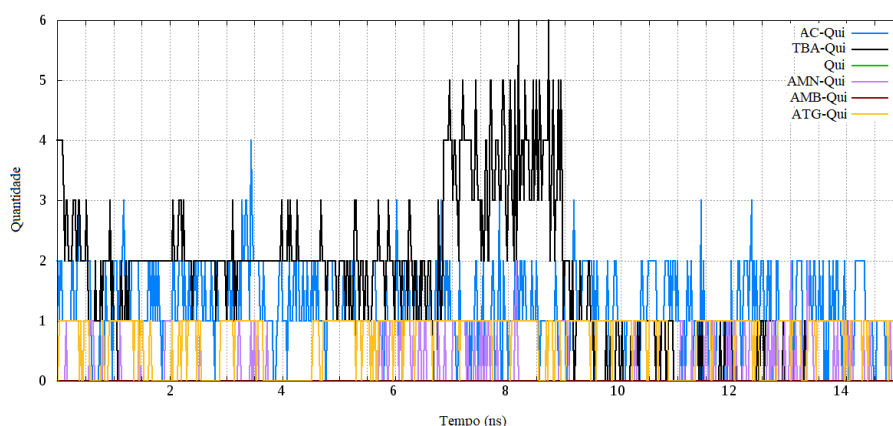


Figura 42. Número de ligação hidrogênio receptor-ligante somente para o substituinte do hexâmero em função do tempo

No caso do AC-Qui-VWD2-VWD3, pode-se notar que as interações de van der Waals são as grandes responsáveis pela estabilização do complexo, perfazendo 29% do $\Delta G_{(ligação-MM/PBSA)}$ quando se considera os valores negativos (TABELA 4), refletido no perfil do *RMSD* do AC-Qui (FIGURA 43) ao passo que para o TBA-Qui esse tipo de interação corresponde a 13% do total das energias estabilizantes (TABELA 4).

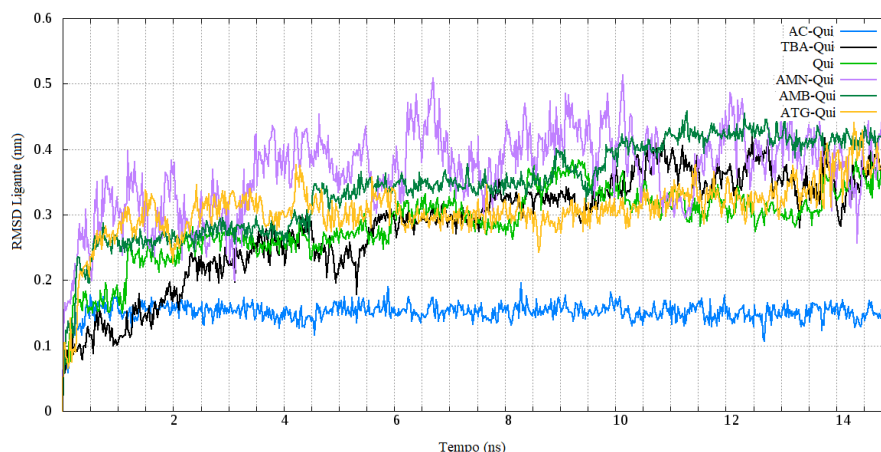


Figura 43. *RMSD* para os ligantes em função do tempo

Outro ponto que merece destaque é o valor de $\Delta G_{(ligação-MM/PBSA)}$ para o ligante Qui, terceiro melhor resultado. Ele representa o polímero quitosana sem as modificações pelo acréscimo de substituintes tiolados em sua estrutura, em que o $\Delta G_{(ligação-MM/PBSA)}$ sugere que o ligante Qui, considerando as interações não covalentes entre receptor-ligante supera, em afinidade, a maioria dos ligantes tiolados. O melhor desempenho do ligante AMN-Qui em relação ao AMB-Qui, pode-se atribuir ao fato do substituinte N,6-ácido mercaptanonicotínico-quitosana (AMN-Qui) ser um melhor acceptor para ligação hidrogênio que o N,4-ácido mercaptanobenzoico-quitosana (AMB-Qui). Quando se compara os valores de *RMSF* da proteína na ausência e em interação com os ligantes (FIGURA 44), observa-se uma estabilização geral na estrutura da proteína induzida pela presença dos ligantes. Havendo uma maior estabilização quando se considera os três ligantes com melhores energias de interação, principalmente na faixa de resíduos a partir de 170.

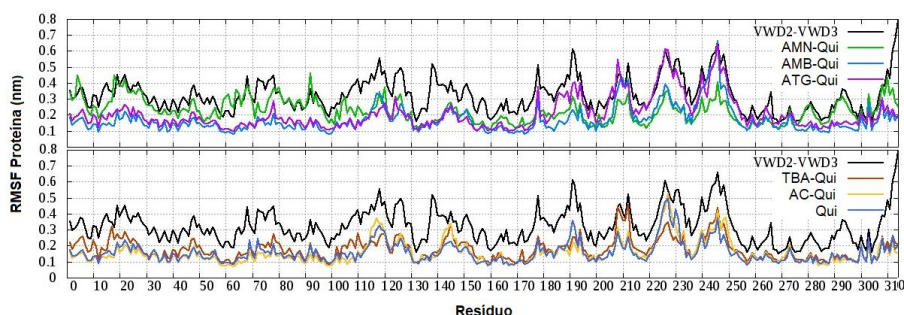


Figura 44. Comparação dos *RMSF* do modelo VWD2-VWD3 sem (preto) e com os diferentes ligantes (coloridos)

5 CONCLUSÃO

Facilitar a aceitação das terapias medicamentosas por parte dos pacientes é um imperativo que se pode alcançar através de pesquisas com quitosanas mucoadesivas. Esse processo passa, inevitavelmente, por considerações sobre aspectos estruturais de mucinas que infelizmente não são experimentalmente conhecidos. As técnicas de modelagem estrutural são ferramentas valiosas que podem contornar essa limitação. Neste trabalho, obteve-se por meio da modelagem comparativa e *threading* um modelo parcial da MUC2 que permitiu extrair informações sobre interações não ligadas com seis hexâmeros de quitosana, cinco dos quais modificados com substituintes tiolados.

A técnica *MMPBSA* para o cálculo das energias livres de interação complexo-ligante se mostrou mais eficaz na diferenciação relativa para essas energias entre os seis ligantes estudados. Os resultados chamam a atenção para a alteração na ordem de mucoadesão quando se considera apenas as interações não ligadas entre o modelo de MUC2 e os ligantes em relação aos dados experimentais obtidos por MUELLER, VERROKEN e IQBAL *et al.* (2011), em especial para a AC-Qui e Qui que figuraram como primeiro e terceiro em $\Delta G_{(ligação-MM/PBSA)}$, respectivamente.

5.1 PERSPECTIVAS

Como a partir de métodos computacionais, onde são utilizadas as equações da mecânica clássica, não é possível prever quebra nem formação de ligações, como perspectivas deste trabalho, quadros de menores energias serão extraídos e, através do uso do programa *MOPAC2016*, tentar-se-á propor mecanismos de reação entre aminoácidos do receptor MUC2 e os ligantes tiolados. Desse modo, espera-se encontrar uma forma de trabalhar com esses métodos, de forma a permitir determinar a influência na afinidade receptor-ligante das interações por meio de ligações dissulfeto, de grande importância no aumento da mucoadesão desses ligantes em relação a quitosana não modificada.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHAM, M. J.; MURTOLA, T.; SCHULZ, R.; PALL, S.; SMITH, J. C.; HESS, B.; LINDAHL, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, p. 19-25, 2015.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Biologia molecular da célula*. 5ª Ed. Porto Alegre: Artmed Editora, 2009. 1294 p.

AMBORT, D.; VAN DER POST, J.; JOHANSSON, M. E. V.; MACKENZIE, J.; THOMSSON, E.; KRENGEL, U.; HANSSON, G. C. Function of the CysD domain of the gel-forming MUC2 mucin. *Biochemical Journal*, **15**, p.61-67, 2011.

ATKINS, P.; JONES, L. *Princípios de química*. 3ª Ed. Porto Alegre: Bookman, 2006. 965 p.

BADHE, R. V.; NANDA, R. K.; CHEJARA, D. R.; CHOONARA, Y. E.; KUMAR P.; TOIT, L. C.; PILLAY, V. Microwave-assisted facile synthesis of a new tri-block chitosan conjugate with improved mucoadhesion. *Carbohydrate Polymers*, **130**, p. 213-221, 2015.

BANSIL, R.; TURNER, B. S. Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science*, **11**, p. 164-170, 2006.

BONTHRON, D. T.; HANDIN, R. I.; KAUFMAN, R. J.; WASLEY, L. C.; ORR, E. C.; MITSOCK, L. M.; EWENSTEIN, B.; LOSCALZO, J.; GINSBURG, D.; ORKIN, S. H. Structure of pre-pro-von Willebrand factor and its expression in heterologous cells. *Nature*, **324**, p. 270 - 273, 1986.

CARVALHO, F. C.; CHORILLI, M.; GREMIÃO, M. P. D. Plataformas bio(muco) adesivas poliméricas baseadas em nanotecnologia para liberação controlada defármacos - propriedades, metodologias e aplicações. *Polímeros*, **24**, n. 2, p. 203-213, 2014.

CHIVANGKUL, T.; PENGPRECHA, S.; PADUNGROS, P.; SIRALEARTMUKUL, K.; PRASONGSUK, S.; MUANGSIN, N. Enhanced water-solubility and mucoadhesion of N,N,N-trimethyl-N-gluconate-N-homocysteine thiolactone chitosan. *Carbohydrate Polymers*, **108**, p. 224–231, 2014.

EIDHAMMER, I.; JONASSEN, I.; TAYLOR, W. Protein bioinformatics: an algorithmic approach to sequence and structure analysis. West Sussex: Wiley, 2004. 355 p.

FOX, N. K.; BRENNER, S. E.; CHANDONIA, J. The value of protein structure classification information-Surveying the scientific literature. *PROTEINS*, **83**, p. 1-14, 2015.

GENHEDEN, S.; RYDE U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*, n. 5, p. 449-461, 2015.

GINALSKI, K. Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology*, **16**, n. 2, p. 172-177, 2006.

HANWELL, M. D.; CURTIS, D. E.; LONIE, D. C.; VANDERMEERSCH, T.; ZUREK, E.; HUTCHISON, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, **4**, p. 1-17, 2012.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **14**, p. 33-38, 1996.

JOHANSSON, M. E. V.; LARSSON, J. M. H.; HANSSON, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. *PNAS*, **108**, p. 4659-4665, 2011.

JORIEUX, S.; FRESSINAUD, E.; GOUDEMANT, J.; GAUCHER, C.; MEYER, D.; MAZURIER, C. Conformational changes in the D' domain of von Willebrand factor induced by CYS 25 and CYS 95 mutations lead to factor VIII binding defect and multimeric impairment. *BLOOD*, **95**, n.10, p. 3139-3145, 2000.

KÄLLBERG, M.; WANG, H.; WANG, S.; PENG, J.; WANG, Z.; LU, H.; XU, J. Template-based protein structure modeling using the *RaptorX* web server. *Nature Protocols*, **7**, p. 1511–1522, 2012.

KUMAR, S.; VIMAL, A.; KUMAR, A. Why Chitosan? From properties to perspective of mucosal drugdelivery. *International Journal of Biological Macromolecules*, **91**, p. 615-622, 2016.

LANCIA, G.; ISTRAIL, S. Protein Structure Comparison: Algorithms and Applications, *Mathematical Methods for Protein Structure Analysis and Design*, p. 1-33, 2003.

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

LEWARS, E. G. *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*. London New York: Springe, 2011. 680 p.

LÜTHY, R.; BOWIE, J.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature*, 356, p. 83-85, 1992.

MORGON, Nelson H.; COUTINHO, K. *Métodos de Química Teórica e Modelagem Molecular*. São Paulo: Editora Livraria da Física, 2007. 539 p.

MORRIS, G. M.; HUEY, R.; LINDSTROM, W.; SANNER, M. F.; BELEW, R. K.; GOODSELL, D. S.; OLSON, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry*, **16**, 2785-9, 2009.

MUELLER, C.; VERROKEN, A.; IQBAL, J.; BERNKOP-SCHNUERCH, A. Thiolated Chitosans: In Vitro Comparison of Mucoadhesive Properties. *Journal of Applied Polymer Science*, 124, p. 5046-5055, 2011.

O'LEARY, J. M.; HAMILTON, J.M.; M. DEANE, C. M.; VALEYEV, N. V.; SANDELL, L. J.; DOWNING, A. K. Solution Structure and Dynamics of a Prototypical Chordin-like Cysteine-rich Repeat (von Willebrand Factor Type C Module) from Collagen IIA. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, **279**, p.53857-53866, 2004.

PAISSONI, C., SPILIOTOPOULOS, D., MUSCO, G. SPITALERI, A. GMXPBSA 2.0: A GROMACS tool to perform MM/PBSA and computational alanine scanning. *Comput. Phys. Commun*, **186**, p. 105-107, 2015.

PAISSONI, C., SPILIOTOPOULOS, D., MUSCO, G. SPITALERI, A. GMXPBSA 2.0: A GROMACS tool to perform MM/PBSA and computational alanine scanning. *Comput. Phys. Commun*, **185**, p. 2920-2929, 2014.

PELASEYED, T.; BERGSTRÖM, J. H.; GUSTAFSSON, J. K. ERMUND, A.; BIRCHENOUGH, G. M. H.; SCHÜTTE, A.; POST, S. V.; SVENSSON, F.; RODRIGUEZ-PINEIRO, A. N.; NYSTRÖM, E. E. L.; WISING, C.; JOHANSSON, M. E. V.; HANSSON, G. C. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunological Reviews*, **260**, p. 8-20, 2014.

PETTERSEN, E. F.; GODDARD, T. D.; HUANG, C. C.; COUCH, G. S.; GREENBLATT, D. M.; MENG, E. C.; FERRIN, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, n.13, p. 1605-1612, 2004.

RIBEIRO, A. A. S. T.; HORTA, B. A. C.; ALENCASTRO, R. B. MKTOP: a Program for Automatic Construction of Molecular Topologies. *J. Braz. Chem. Soc.* **19**, n. 7, p. 1433-1435, 2008.

RINO, J. P.; da COSTA, B. V. ABC da simulação computacional. São Paulo: Livraria da Física, 2013, 191 p.

ROBERTSON, M. J.; TIRADO-RIVES, J.; JORGENSEN, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* 11, n.7, p 3499–3509, 2015.

RWEI, S.; LIEN C. Synthesis and viscoelastic characterization of sulfonated chitosan solutions. *Colloid Polym Sci*, **292**, p. 785–795. 2014.

SAXENA, A.; SANGWAN, R. S. Fundamentals of Homology Modeling Steps and Comparison among Important Bioinformatics Tools: An Overview. *Science International*, **1**, n. 7, p. 237-252, 2013.

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

SCHLICK, T. Molecular modeling and simulation: an interdisciplinary guide. New York: Springer, 2006. 467 p.

SHRIVER, D. F.; ATKINS, P. W.; OVERTON, T. L.; ROURKE, J. P.; WELLER, M. T.; ARMSTRONG. SHRIVER e ATKINS: química inorgânica. 4ª Ed. Porto Alegre: Bookman, 2008. 847 p.

SIEVERS, F.; WILM, A.; DINEEN, D. G.; GIBSON, T. J.; KARPLUS, K.; LI, W. LOPEZ, R.; MCWILLIAM, H.; REMMERT, M.; SÖDING, J.; THOMPSON, J. D.; HIGGINS, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, n. 539 p. 1-6, 2011.

da SILVA, A. W. S.; VRANKEN, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, **5**, p. 1-8, 2012.

STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model*, **19**, p. 1-32, 2013.

SUDHEESH K. S.; AJAY K. M.; OMOTAYO A. A.; BHEKIE B. M. Chitosan-based nanomaterials: A state-of-the-art review. *International Journal of Biological Macromolecules*, **59**, p.46-58, 2013.

TEGGE, A. N.; WANG, Z.; CHENG, J. Integrative protein fold recognition by alignments and machine learning. In: RANGWALA, H.; KARYPIS, G. (org.). *Introduction to Protein Structure Prediction: Methods and Algorithms*. New Jersey: WILEY, 2010, 500p.

TROTT, O.; OLSON, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of *docking* with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry*, **31**, p. 455-461, 2009.

VERLI, H. *Bioinformática: da Biologia à Flexibilidade Molecular*. São Paulo: SBBq, 2014, 282 p.

WU, S. ZHANG, Y. *LOMETS*: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research*, **35**, p. 3375-3382, 2007.

XIONG, J. *Essential Bioinformatics*. New York: United States of America by Cambridge University Press, 2006. 362 p.

XU, D.; JAROSZEWSKI, L.; LI, Z.; GODZIK, A. *AIDA*: ab initio domain assembly for automated multi-domain protein structure prediction and domain–domain interaction prediction. *Bioinformatics*, **31**, n. 13, p. 2098-2105, 2015.

APÊNDICE A

PASSOS SEGUIDOS NA EXECUÇÃO DO TRABALHO PARA DINÂMICA MOLECULAR DO RECEPTOR E CÁLCULO DE ENERGIA DE COMPLEXAÇÃO POR MEIO DO PROGRAMA GMXPBSA

DINÂMICA MOLECULAR NO GROMACS-5.x.x

O primeiro passo é dispor de um arquivo com extensão .pdb de uma proteína, esse tipo de arquivo contém as coordenadas 3D de estruturas de moléculas ou macromoléculas, como exemplo aqui, o modelo VWD2-VWD3.pdb obtido da modelagem de parte da sequência primária da mucina 2. Porém, pode-se obter arquivos .pdb de proteínas com estruturas resolvidas experimentalmente com facilidade no Protein DataBank (<http://www.rcsb.org/pdb/home/home.do>), ou então, buscar qualquer molécula no banco de dados ChemSpider (<http://www.chemspider.com/>), baixar uma molécula no formato .mol e converter para .pdb utilizando o programa OpenBabel (http://openbabel.org/wiki/Main_Page). Exemplificando, baixe a molécula 3D 258.mol (butanol no ChemSpider sob o código de identificação 258), utilize o seguinte comando no terminal do linux para converter de .mol para .pdb, estando na pasta onde se encontra o arquivo, é claro:

```
$ babel -imol 258.mol -opdb butanol.pdb
```

Porém, quando não se objetiva dinâmicas apenas de proteínas, é necessário parametrizar a molécula para o campo de forças desejado. Ver-se-á como parametrizar uma molécula em outro momento. Bem, dispondo do arquivo .pdb na pasta onde irá rodar a simulação, ative o GROMACS com o seguinte comando:

```
$ source usr/local/gromacs/bin/GMXR
```

aperte ENTER, isso porque se instalou o GROMACS no diretório e subdiretório usr/local).

Com o GROMACS ativado e o arquivo.pdb na pasta, entre com o seguinte comando:

```
$ gmx pdb2gmx -f VWD2-VWD3.pdb -o sistema.gro -p sistema.top
```

Como resultado desse comando, o terminal irá exibir uma lista de **Campo de Forças**, neste momento digite o número referente ao Campo de Forças de sua preferência e, em seguida ENTER, no caso específico da desta simulação a opção do Campo de Forças foi o OPLS-AA de número 15. Outra lista será exibida no terminal, porém, com os **modelos de água** que deseja usar em sua simulação, aqui, a opção foi pelo modelo TIP3P cuja opção de escolha é o número 3. Uma observação importante, use os mesmos nomes para os arquivos de saída que estão sendo utilizando aqui, salvo o arquivo.pdb inicial que deve ser idêntico ao nome do arquivo da sua proteína, nesse exemplo, o arquivo VWD2-VWD3.pdb, padronizando os nomes dos *input* e *output*, facilita muito a realização das simulações e com o tempo você vai se familiarizando e conhecendo o que cada um representa em cada etapa do processo.

Para versões mais antigas do GROMACS não é preciso incluir **gmx** na frente dos comandos e para outras os comandos são diferentes.

O programa GMX PDB2GMX converter o arquivo.pdb para o arquivo de entrada (*input*) do GROMACS, os arquivos de extensão .gro com as coordenadas do sistema, além de gerar o arquivo contendo a **topologia** da proteína com o nome de saída (*output*) sistema.top. Para entender o que cada comando faz digite no terminal o nome do programa seguido de -h, exemplo, para ter uma descrição do programa GMX PDB2GMX, digite

```
$ gmx pdb2gmx -h
```

aperte ENTER. Assim você vai ter acesso a cada palavra-chave do comando, bem como sua função.

Dando continuidade à preparação dos arquivos para a realização da dinâmica, entre com o seguinte comando:

```
$ gmx editconf -bt octahedron -princ -f sistema.gro -o sistema_box.gro -c -d 1.0
```

Será exibido um índice com as opções para a orientação do sistema, escolha o soluto, ou seja, a proteína apertando 0 ou 1, em seguida ENTER. Com esse comando cria-se uma caixa octaédrica (-bt octahedron), alinhada com os eixos principais (-princ), inserindo a molécula (-f sistema.gro) no centro da caixa (-c),

mantendo a molécula longe das paredes da caixa a uma distância de 1 nanômetro (nm) (-d 1.0).

Agora entre com o comando abaixo:

```
$ gmx solvate -cp sistema_box.gro -cs spc216.gro -o sistema_boxwater.gro -p sistema.top
```

Com o comando GMX SOLVATE adiciona-se moléculas de água no interior da caixa (-cp sistema_box.gro -cs spc216.gro), gerando como saída a proteína solvatada (-o sistema_boxwater.gro) e a topologia da nova configuração do sistema (-p sistema.top).

Agora entre com o comando:

```
$ gmx make_ndx -f sistema_boxwater.gro
```

Como resultado será exibido uma lista com quais elementos devem aparecer no arquivo index.ndx, selecione o número 0 e, em seguida aperte ENTER. Esse comando gera um índice das moléculas do sistema, requerido por muitos programas no GROMACS.

MINIMIZAÇÃO DE ENERGIA *STEEPEST DESCENT* COM RESTRIÇÃO DE POSIÇÃO

Agora fique atento as informações mostradas no terminal depois da execução do programa GMX GROMPP, pois se o sistema não estiver com suas cargas equilibradas, você saberá nesse momento, o GMX GROMPP avisa se houver necessidade de adicionar íons para balancear as cargas:

```
$ gmx grompp -f stpr.mdp -c sistema_boxwater.gro -p sistema.top -o sistema_stpr.tpr -maxwarn 40 -n
```

O programa GMX GROMPP gera o arquivo binário (-o sistema_stpr.tpr) com todas as informações necessárias para rodar as simulações, as vezes o programa faz notificações sobre os parâmetros da minimização de energia *steepest descent* com restrição de posição (-f stpr.mdp), se as notificações forem consideráveis ele aponta erro e não gera o binário, se seus parâmetros estiverem corretos você pode ignorar os avisos (-maxwarn 40), neste momento entra-se também com o arquivo de index

gerado com o programa GMX MAKE_NDX (-n).

O próximo comando é apenas para aqueles sistemas cuja carga total não seja zero, o primeiro é para sistemas que estejam com a carga total negativa e o segundo para carga total positiva. Caso seu sistema já esteja neutro, vá direto para o programa GMX MDRUN.

```
$ gmx genion -s sistema_stpr.tpr -o sistema_ions.gro -pname NA -np 3 -pq 1 -p
sistema.top
```

```
$ gmx genion -s sistema_stpr.tpr -o sistema_ions.gro -nname CL -nn 3 -nq -1 -p
sistema.top
```

As diferenças importantes entre os dois comandos acima estão na indicação do íon sódio (-pname NA) sua carga (-pq 1) e o número de íons necessários, neste caso específico três (-np 3), mas vai depender do módulo da carga negativa do sistema. No segundo, temos as indicações do íon cloreto (-nname CL), número da carga do cloreto (-nq -1) e do número de cloretos necessários para neutralizar o sistema (-nn 3).

Realizado a neutralização do sistema, entre com os comandos:

```
$ gmx grompp -f stpr.mdp -c sistema_ions.gro -p sistema.top -o sistema_stpr.tpr
```

Caso seja necessário entre com a palavra-chave -maxwarn 40 -n.

```
$ gmx mdrun -deffnm sistema_stpr -v -g stpr>&stpr.job&
```

O programa GMX GROMPP você já conhece, já o GMX MDRUN é o que vai rodar sua minimização de energia. Para ver o programa trabalhando execute o comando:

```
$ tail -f stpr.job &
```

Está etapa é relativamente rápida, leva em torno de 15 a 30 minutos.

MINIMIZAÇÃO DE ENERGIA *STEEPEST DESCENT* SEM RESTRIÇÃO DE POSIÇÃO

Essas etapas de minimização e realização da dinâmica propriamente dita segue sempre os mesmos passos, lembrando que para cada etapa é necessário um arquivo.mdp com os parâmetros necessários para a instrução do GROMACS (<http://manual.gromacs.org/documentation/2018/user-guide/mdp-options.html>) ou (<http://www.gromacs.org/@api/deki/files/198/=gmx-tutorial.pdf>).

```
$ gmx grompp -f st.mdp -c sistema_stpr.gro -p sistema.top -o sistema_st.tpr
```

```
$ gmx mdrun -deffnm sistema_st -v -g st>&st.job&
```

Para ver o programa trabalhando execute o comando:

```
$ tail -f st.job &
```

DINÂMICA DE ÁGUA

Com esse próximo passo no processo de minimização de energia do sistema a ser simulado, far-se-á com que as moléculas de água se movimentem no sistema, porém o soluto permanece parado, isso permite um ajuste das moléculas do solvente em torno da proteína. Para realizar a dinâmica de água, entre com os seguintes comandos:

```
$ gmx grompp -f pr.mdp -c sistema_st.gro -p sistema.top -o sistema_pr.tpr -maxwarn 40
```

```
$ gmx mdrun -deffnm sistema_pr -v -g pr>&pr.job&
```

Para ver o programa trabalhando execute o comando:

```
$ tail -f pr.job &
```

Esta etapa já é mais demorada, pode levar de 5 a 12 horas a depender da capacidade de processamento do seu computador.

DINÂMICA SOLTA

Os passos seguidos até aqui serviram para remover possíveis tensões no sistema de simulação, sem eles sua simulação pode apresentar problemas e até mesmo o rompimento de ligações o que se costuma chamar de explosão do sistema. A partir

daqui sua dinâmica terá efetivamente início.

Entre com os comandos:

```
$ gmx grompp -f md.mdp -c sistema_pr.gro -p sistema.top -o sistema_md.tpr -maxwarn 40
```

```
$ gmx mdrun -deffnm sistema_md -v -g md>&md.job&
```

```
$ tail -f md.job &
```

REINICIAR DINÂMICA INTERROMPIDA

Por diversas razões sua dinâmica pode ser interrompida, ou até mesmo você desejar interrompê-la por algum motivo, caso isso aconteça, você pode recomeçar a simulação de onde ela tenha parado, sem a necessidade de recalculá-la tudo novamente, para isso, basta entrar com o comando abaixo:

```
$ gmx mdrun -cpi sistema_md.cpt -deffnm sistema_md -v -g md>&md.job&
```

A palavra-chave `-cpi`, seguida do arquivo `sistema_md.cpt` sinaliza ao programa GMX MDRUN que retome a simulação de um ponto um pouco antes do momento em que ela fora interrompida, o arquivo `.cpt` é gerado logo após o início da simulação, onde fica gravado as instruções e informações para o caso de a dinâmica ser interrompida inadvertidamente, caso esse arquivo esteja corrompido, existe o arquivo `_prev.cpt` que serve para o mesmo fim, então se você não conseguir retomar sua dinâmica, substitua o `sistema_md.cpt` pelo arquivo `sistema_md_prev.cpt`.

EXTENDENDO A DINÂMICA

Os comandos adiante nos próximos dois itens só servem para o caso de precisar estender sua simulação, se não for esse o seu caso, pule-o, vá direto para o item ANÁLISE DOS RESULTADOS.

Caso tenha terminado a sua dinâmica, porém acredita que poderia tê-la feito um pouco mais longa, é possível continuar a partir de onde parou com apenas dois comandos, por exemplo, foi preparado inicialmente o arquivo `md.mdp` para rodar 30 ns, ao fim da dinâmica, deseja-se aumentá-la em 10 ns, de modo que se chegue a 40 ns. Para isso, basta executar os seguintes comandos:

```
$ gmx convert-tpr -s sistema_md.tpr -f sistema_md.cpt -extend 10000 -o
sistema_md2.tpr
```

```
$ gmx mdrun -deffnm sistema_md2 -v -g md2>&md2.job&
```

O valor 10000 no primeiro comando corresponde aos 10 ns que se deseja aumentar, porém em picossegundos (ps), se quiser aumentar mais 5, 3, 20 enfim, não importa a quantidade, basta digitar o valor em ps no lugar dos 10000 do primeiro comando. Lembre-se, caso precise aumentar a dinâmica mais de uma vez, vá alterando o input sistema_md.cpt para sistema_md2.cpt e o output de sistema_md2.tpr para sistema_md3.tpr e assim sucessivamente. Caso o seu arquivo sistema_md.cpt esteja corrompido por alguma razão, substitua-o pelo arquivo sistema_md_prev.cpt.

JUNTANDO OS ARQUIVOS GERADOS NA DINÂMICA INICIAL E APÓS A DINÂMICA EXTENDIDA

Findado a dinâmica estendendo a simulação inicial, como nesse exemplo, passa-se a ter os arquivos sistema_md.xtc e sistema_md2.xtc de trajetórias e os arquivos sistema_md.edr e sistema_md2.edr de energias, é preciso, então juntá-los num arquivo único que iram se chamar de sistema_mdt.xtc e sistema_mdt.edr respectivamente. Para isso, usar-se-á os seguintes comandos:

para as trajetórias:

```
$ gmx trjcat -f sistema_md.xtc sistema_md2.xtc -o sistema_mdt.xtc -cat -settime
```

O programa fará uma solicitação, aperte a letra C e dê ENTER para cada início de trajetória.

Verifique se o arquivo.xtc gerado tem as trajetórias correspondentes a soma dos tempos de simulação de cada um dos *inputs*, 40000 ps no exemplo utilizado aqui, faça isso com o comando:

```
$ gmx check -f sistema_mdt.xtc
```

para as energias:

```
$ gmx eneconv -f sistema_md.edr sistema_md2.edr -o sistema_mdt.edr
```

Verifique se o arquivo.edr gerado tem as energias correspondentes a soma dos tempos de simulação de cada um dos inputs, 40000 ps no exemplo utilizado aqui, faça isso com o comando:

```
$ gmx check -e sistema_mdt.edr
```

Caso tenha mais arquivos de trajetória e energias para juntar, basta acrescentá-los na sequência de *input* por exemplo, -f sistema_md.xtc sistema_md2.xtc sistema_md3.xtc ..., valendo o mesmo para os arquivos de energias. Os arquivos sistema_md.tpr e sistema_md2.tpr gerados antes e após a extensão da simulação são os mesmos, por essa razão, não é necessário juntá-los, nas ocasiões em que se precisar deles, pode-se utilizar tanto um quanto o outro indistintamente.

ANÁLISE DOS RESULTADOS

Analisar a dinâmica do seu sistema é um passo de grande importância, afinal é por isso que realizam-se as simulações, tendo sempre em vista as informações que elas podem fornecer, o GROMACS tem uma infinidade de ferramentas de análise, cada uma adquirindo importância a partir do tipo de informação que se deseja extrair do sistema, aqui se mencionará apenas algumas análises básicas que permitem verificar se o sistema entrou em equilíbrio, o que é fundamental para a validade das informações extraídas da simulação.

Antes de entrar na análise da dinâmica propriamente dito, deve-se checar se a proteína e as moléculas do solvente não enxergaram suas próprias imagens durante a simulação, caso isso aconteça a dinâmica estará invalidada. Para checar a distância mínima que uma molécula chegou da sua própria imagem, ou seja, se o tamanho da menor aresta da caixa de simulação respeitou a regra estabelecida na qual diz que a menor aresta deva ser maior ou igual ao comprimento da molécula ou macromolécula mais duas vezes o raio de corte:

Menor aresta \geq comprimento da molécula + 2x raio de corte

Para essa finalidade o GROMACS dispõe de uma ferramenta chamada GMX MINDIST. Para executá-la, utiliza-se os seguintes comandos

```
$ gmx mindist -f sistema_md.xtc -s sistema_md.tpr -od md_mindist.xvg -pi
```

Escolha a opção 1 (protein) no terminal, esse é um processo um pouco demorado. Para analisar o arquivo gerado abra-o com o comando:

```
$ xmgrace md_mindist.xvg&
```

Validado a simulação é hora de analisar os resultados. Para isso é necessário realizar algumas tarefas preliminares. A simulação inclui todo o sistema com íons, água e proteína, porém para a análise do Desvio da Raiz Média Quadrática (RMSD) em inglês, RMSF e DSSP é preciso separar a proteína do restante do sistema, para tanto vamos executar os seguintes comandos:

```
$ gmx trjconv -s sistema_md.tpr -f sistema_md.xtc -o proteina.xtc
```

Escolha a opção 1 (protein) no terminal.

Com esse comando removem-se as coordenadas de águas e íons do arquivo de trajetória sistema_md.xtc que resulta na criação do novo arquivo de trajetória proteina.xtc.

Já com o comando:

```
$ gmx convert-tpr -s sistema_md.tpr -o proteina.tpr
```

Escolha a opção 1 (protein) no terminal.

Removem-se do arquivo sistema_md.tpr as informações sobre águas e íons, gerando então um outro arquivo chamado proteina.tpr.

Para gerar o arquivo proteina.gro com as coordenadas iniciais da trajetória, usa-se o comando:

```
$ gmx trjconv -s proteina.tpr -f proteina.xtc -o proteina.gro -dump 0
```

No terminal escolha a opção 0 ou 1.

A palavra-chave -dump especifica o tempo de simulação em ps para a criação do arquivo proteina.gro, no nosso exemplo no tempo 0 ps. Se você abrir os arquivos proteina.gro proteina.xtc juntos no VMD com o comando:

```
$ vmd proteina.gro proteina.xtc&
```

E visualizar o filme da simulação. Vai perceber que a proteína está transladando e rotacionando, se ela sair pelo lado da caixa então, ai bagunça tudo, o VMD não consegue entender as condições periódicas de contorno, por isso a visualização fica comprometida, com os átomos e ligações se esticando ou aparentemente rompidas, isso também dificulta a análise posterior, portanto é necessário fitar, ou melhor, ajustar a trajetória da proteína, ou seja, remover esses movimentos translacionais e rotacionais, para isso usa-se o comando:

```
$ gmx trjconv -s proteina.tpr -f proteina.xtc -o proteina_fit.xtc -pbc mol -center
```

Escolha no terminal as opções 1 (protein) e, seguidamente 1 (protein) novamente. Com esse comando remove-se as condições periódicas de contorno (-pbc), centraliza-se a molécula na caixa (-center), a partir do seu centro de massa (mol).

Em seguida o comando:

```
$ gmx trjconv -s proteina.tpr -f proteina_fit.xtc -o proteina_fit.xtc -fit progressive
```

Escolha no terminal as opções 1 (protein) em seguida aperte ENTER, escolha 1 (protein) novamente, em seguida aperte ENTER. Com esse comando ajusta-se os quadros (frames) da dinâmica (-fit) ao quadro anterior (progressive)

Você precisa gerar o arquivo `proteina.gro` novamente, porém com a nova trajetória fitada:

```
$ gmx trjconv -s proteina.tpr -f proteina_fit.xtc -o proteina_fit.gro -dump 0
```

Observe agora como ficou. Abrindo no VMD os arquivos `proteina_fit.xtc` e `proteina.gro` com o comando:

```
$ vmd proteina_fit.gro proteina_fit.xtc &
```

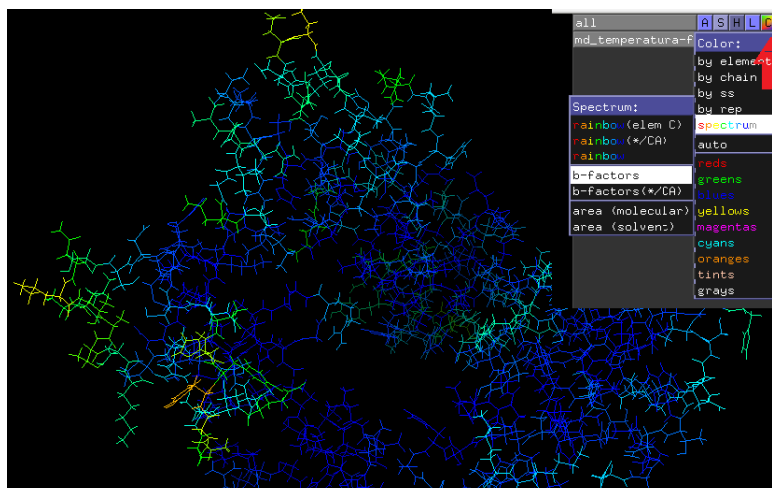
A partir daqui se começa a realizar as análises básicas que podem ser feitas depois de uma simulação de proteína em água.

GERANDO O GRÁFICO RMSF VERSUS RESÍDUO

```
$ gmx rmsf -n index_proteina.ndx -s proteina.gro -f proteina_fit.xtc -ox  
media_proteina.pdb -o rmsf_proteina.svg -oq md_temperatura-factor-residuo.pdb -  
res
```

O comando acima gera tanto o gráfico de RMSF em função dos resíduos (`rmsf_proteina.svg`), quanto a representação do Fator-B (`md_temperatura-factor-residuo.pdb`) Figura 21.

Para visualizar as representações do Fator -B em fita (Figura 21), basta abrir o *output* `md_temperatura-factor-residuo.pdb`, gerado pelo comando anterior, com o programa PYMOL e ir até a opção `b-factors`



Para obter a representação em massa, Também Figura 21, basta entrar, em seguida, com os seguintes comandos, na sequência:

```
PyMOL> spectrum b, blue_white_red, minimum=20, maximum=50
PyMOL> as cartoon
PyMOL> cartoon putty
```

GERANDO O GRÁFICO RMSD VERSUS TEMPO DE SIMULAÇÃO

Inicialmente precisa-se gerar um arquivo de índice contendo os átomos que se irá utilizar através do comando:

```
$ gmx make_ndx -f proteina.tpr -o index_proteina.ndx
```

Escolha no terminal a opção 1 (protein) em seguida aperte ENTER, aperte letra q no teclado e em seguida ENTER novamente.

Com o comando abaixo é possível gerar o arquivo contendo os dados da mudança da estrutura ao longo da simulação:

```
$ gmx rms -n index_proteina.ndx -s proteina.tpr -f proteina_fit.xtc -o
rmsd_proteina.xvg -tu ns
```

```
$ gmx rms -n index_proteina.ndx -s media_proteina.pdb -f proteina_fit.xtc -o
rmsd_prot-X-media-prot.xvg -tu ns
```

Escolha seguidamente 4 (Backbone) e 4 (Backbone), para a o ajuste por mínimos quadrados e medidas do rmsd respectivamente. O arquivo rmsd_proteina.svg pode ser aberto usando o programa GRACE com o comando:

```
$ xmgrace rmsd_proteina.svg&
```

RAIO DE GIRO DA PROTEÍNA

```
$ gmx gyrate -f proteina_fit.xtc -s proteina.tpr -o proteina_raio-de-giro.svg
```

ESTUDO DAS LIGAÇÕES DE HIDROGÊNIO DURANTE A DINÂMICA MOLECULAR

LIGAÇÕES DE HIDROGÊNIO INTRA-CADEIA

```
$ gmx hbond -f proteina_fit.xtc -s proteina.tpr -num proteina_H_intra.svg
```

Escolha sucessivamente as opções 1 (protein) e 1 (protein) no terminal. Abra o arquivo com o comando:

```
$ xmgrace proteina_H_interna.svg
```

LIGAÇÕES DE HIDROGÊNIO ENTRE SOLVENTE E PROTEÍNA

Gerando o arquivo que permite monitorar a quantidade de ligações de hidrogênio entre solvente e proteína ao longo do tempo da dinâmica, utilize o comando:

```
$ gmx hbond -f sistema_mdt.xtc -s sistema_md.tpr -num proteina_H_inter.svg -nonitacc
```

Escolha sucessivamente as opções 1 (Protein) e 13 (SOL) no terminal. Com esse comando exclui-se as amins como aceptoras de hidrogênio (-nonitacc), o que não é uma má consideração pois os Nitrogênios das amidas não são boas aceptoras e as amins estão em sua maioria protonadas em condições fisiológicas.

Abra o *output* gerado com o comando:

```
$ xmgrace sistema_md_H.svg &
```

OBTENÇÃO DE CONFORMAÇÕES MÉDIAS UTILIZANDO O PROGRAMA *GMX CLUSTER*

Os programas de *docking* molecular sofrem de uma limitação que é a de considerar o receptor como uma estrutura rígida durante o atracamento, sabe-se que durante o processo de interação receptor-ligante ambas as estruturas exploram um amplo espaço conformacional, exibindo seus respectivos graus de liberdade, a fim de encontrarem a complementaridade conformacional que minimize as tensões e maximizem as interações. Porém, uma macromolécula como uma proteína tem muitos graus de liberdade disponíveis, tornando essa busca inviável em termos da técnica de *docking* molecular. Um modo de contornar esse problema é utilizando os clusters, que podem ser obtidos seguindo os passos abaixo:

RMSD TODOSXTODOS

```
gmx rms -f proteina_fit.xtc -f2 proteina_fit.xtc -s proteina.tpr -m proteina_RMSD-  
matrix.xpm -b 35000
```

Escolha seguidamente 1 (Protein) e 1 (Protein), para a o ajuste por mínimos quadrados e medidas do rmsd respectivamente.

```
gmx xpm2ps -f proteina_RMSD-matrix.xpm -o proteina_RMSD-matrix.eps -rainbow  
blue
```

```
ps2pdf proteina_RMSD-matrix.eps proteina_RMSD-matrix.pdf
```

Para obter as várias conformações médias a partir de uma simulação é necessário primeiramente preparar alguns arquivos a serem utilizados como entrada no programa *gmx cluster* um deles é a matriz de RMSD que já foi gerada anteriormente

```
gmx cluster -f proteina_fit.xtc -s proteina.tpr -dm proteina_RMSD-matrix.xpm -dist  
proteina_rmsd-distribuição.svg -o proteina_cluster.xpm -sz proteina_cluster-  
tamanho.svg -tr proteina_transição.xpm -ntr proteina_transição.svg -clid  
proteina_cluster-Xtempo.svg -cl proteina_clusters.pdb -cutoff 0.25 -method gromos -  
b 35000
```

Escolha seguidamente 1 (Protein) e 1 (Protein).

RODANDO O PROGRAMA GMXPBSA

Inicialmente foi necessário baixar o programa, a versão GMXPBSA-2.1.2 é a recomendada pelos criadores do programa, pois as versões anteriores estavam apresentando erro com o uso do Gromacs-5.x.x. Para usar o programa é preciso ter instalado o GROMACS e o programa APBS, pois o GMXPBSA chama sub-rotinas daqueles dois programas. Uma vez os três programas devidamente instalados, é preciso criar o arquivo chamado INPUT.dat, neste arquivo é colocado as palavras-chave a partir das quais o programa executará o cálculo da variação da energia livre de Gibbs. Abaixo um exemplo do arquivo INPUT.dat usado no trabalho realizado com os modelos de MUC2 e quitosanas:

```
# CONTEÚDO DO ARQUIVO INPUT.dat

root    QUI-MMPBSA # este termo refere-se a pasta que deve conter os arquivos .xtc; .tpr e .ndx
name_xtc  proteina-ligante_fit # indica qual arquivo .xtc deve ser usado pelo GMXPBSA
name_tpr  proteina-ligante # indica qual arquivo .tpr deve ser usado pelo GMXPBSA
complex  Protein_LIG # indica o nome do complexo receptor-ligante no index.ndx
receptor  Protein # indica o nome do receptor no index.ndx
ligand    LIG # indica o nome do ligante no index.ndx
ffield    15 # campo de força usado na dinâmica, o número 15 indica o OPLS no PDB2GMX
use_topology  y # indica que se fará uso de topologias
ligand.itp  # indica a topologia do ligante
itp_ligand  QUI-opt_GMX_OPLS.itp # nome do arquivo da topologia do ligante
itp_receptor  topol.top # nome da topologia do receptor, deve ser renomeado como topol.top
multitry   n # indica que apenas um sistema será tratado
cluster    n # indica o uso ou não de cluster, neste caso a opção é n de negativo
run        1 # indica quantos cálculos será executado para cada sistema

Gpath     /usr/local/gromacs/bin/ # caminho onde os binários do GROMACS serão encontrados
Apath     /home/ocs/bin/ # caminho onde os binários do APBS serão encontrados
Cpath     /opt/apbs-1.3-source/tools/manip/ # onde o binário do coulomb será encontrado
coulomb    coul # indica que a energia coulômbica será calculada pelo APBS
pdf       n # opção não para output em pdf
```

```
min      n # opção não para cálculo de minimização de energia para cada frame
skip     7 # indica que se deve pegar 1 frame e pular 7 ao longo de toda a trajetória
total_number_of_jobs 107 # número total de frames da trajetória dividido por sete (skip=7)
mnp      1 # define o número máximo de processadores usado pelo GMXPBSA
Recoverjobs y # se um cálculo PBSA apresentar erro, ele é calculado novamente
temp     310 # indica a temperatura para o cálculo PBSA pelo APBS
```

O arquivo INPUT.dat precisa estar dentro do mesmo diretório que contenha a pasta com os arquivos .xtc, .tpr, .ndx, .itp e .top.

Por exemplo, para determinar a energia livre da dinâmica QUI-quitosana e receptor por *MMPBSA*, precisou-se gerar o index.ndx da proteína do ligante e do complexo proteína-ligante, o index deve ter apenas a descrição desses três elementos, se for o caso abra o index.ndx usando um editor de texto e apague outros elementos que estiverem lá. Protein, LIG e Protein_LIG foi como o programa MAKE_NDX indicou o nome de cada elemento receptor, ligante e complexo respectivamente, são esses nomes que devem constar no arquivo INPUT.dat na frente dos termos receptor, ligand e complex, como no exemplo acima.

antes de rodar o comando gmxpbsa0, rodar os comandos:

```
$ OMP_NUM_THREADS=1
```

```
$ export LC_NUMERIC="en_US.UTF-8"
```

APÊNDICE B

DETALHES DOS CONTATOS ENTRE AMINOÁCIDOS DO MODELO VWD2-VWD3 E LIGANTES RESULTANTES DO DOCKING. AC-QUI, AMB-QUI, AMN-QUI, ATG-QUI, QUI E TBA-QUI RESPECTIVAMENTE.

