



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Raphael Vinicius Rodrigues Dias

Avaliação computacional de mecanismos de quinases e proteínas relacionadas

São José do Rio Preto
2018

Raphael Vinicius Rodrigues Dias

Avaliação computacional de mecanismos de quinases e proteínas relacionadas

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto

Financiadora: FAPESP – Processo 2016/08753-6

Orientador: Prof. Dr. Leandro Cristante de Oliveira

Co-orientador: Prof. Dr. Fernando Alves de Melo

São José do Rio Preto
2018

Dias, Raphael Vinicius Rodrigues.

Avaliação computacional de mecanismos de quinases e proteínas relacionadas / Raphael Vinicius Rodrigues Dias. -- São José do Rio Preto, 2018
100 f. : il.

Orientador: Leandro Cristante de Oliveira

Coorientador: Fernando Alves de Melo

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Biofísica. 3. Fosfotransferases. 4. Dinâmica molecular. 5. Proteínas – Análise. 6. Conformação de proteína alfa-hélice. 7. Simulação de docking molecular. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas.
II. Título.

CDU – 577.3

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Raphael Vinicius Rodrigues Dias

Avaliação computacional de mecanismos de quinases e proteínas relacionadas

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto

Financiadora: FAPESP – Processo 2016/08753-6

Comissão Examinadora

Prof. Dr. Leandro Cristante de Oliveira
Unesp - São José do Rio Preto – SP
Orientador

Profa. Dra. Ana Carolina Migliorini Figueira
Lnbio – Campinas – SP

Prof. Dr. Jorge Chahine
Unesp – São José do Rio Preto – SP

São José do Rio Preto
26 de abril de 2018

*Se eu pudesse, faria alguma coisa verdadeiramente importante,
que definitivamente ficaria para o futuro. E assim, se eu desaparecer,
eu acredito que tudo que eu consegui iria permanecer.*

É isso que significa viver, não é mesmo?

- Grovyle

.



Agradecimentos

A realização desta dissertação contou com importantes contribuições e incentivos, sem os quais não teria se tornado uma realidade e aos quais sou eternamente grato.

Ao Prof. Dr. Leandro Cristante de Oliveira, o melhor orientador que uma pessoa poderia ter. Agradeço por todo o apoio, por sua disponibilidade (todas as tardes, em que eu aparecia na sua sala e ficava horas tirando dúvidas, inclusive nas sextas), pela confiança e por todo o aprendizado que me proporcionou.

Ao Prof. Dr. Fernando Alves de Melo, que aceitou um agregado ao seu grupo de pesquisa. Obrigado pela disponibilidade para discutir qualquer assunto, pela confiança, pelo incentivo e por todo ensinamento.

A Carolina Tatiani. Muitíssimo obrigado pelas palavras de apoio nos momentos difíceis. Você foi um dos pontos mais essenciais para tudo dar certo. Cada conversa, sobre os mais variados assuntos, cada conselho e por sua disponibilidade para todos os momentos. Inclusive, das mais de 50 viagens que combinamos, só algumas já se cumpriram, ou seja, ainda teremos muito tempo viajando.

A Karoline Sanches (Karolinda). Obrigado por tudo! Pela confiança depositada na pesquisa e por cada discussão (e foram muitas) sobre cada tópico de nosso trabalho. Muito obrigado por sua amizade, fortalecida a cada dia que passa. É incrível sua capacidade de me animar, mesmo nos momentos mais difíceis. Dividir cada momento do mestrado contigo, inclusive a pesquisa, foi uma honra!

Ao Paulo Henrique. Mesmo de longe, você foi um dos pontos essenciais para esse trabalho. Meu muito obrigado por cada momento juntos, por cada discussão, por me privilegiar em aprender sempre mais sobre a ciência e fazer eu me apaixonar cada vez mais pela Física. Obrigado também por cada festa e Vila Dionísio juntos, aconteceram tantas coisas que seriam impossíveis citar todos. São lembranças que ficaram para sempre em minha memória!

Aos amigos “os menino” Renan, Naza, Ingrid, João, Thalita, Coragem, Acerola, Pamonha e Zezinho que participaram comigo deste trajeto! Muito obrigado pela amizade e companheirismo. Cada momento com vocês será eternamente lembrado. Destaco aqui o “bar do Helinho”, que noite foi aquela Acerola?

A minha família pelo amor, carinho e compreensão em toda essa jornada de mestrado. Vocês foram as peças-chaves para que tudo desse certo!

Agradecimentos finais à FAPESP, CAPES e CNPq, pelo auxílio financeiro.

Resumo

A proteína *Carboxyl-terminal Src Kinase*, Csk, é uma proteína tirosina quinase com função de reguladora essencial para as proteínas da família das Src quinases (SFKs). Seu mal funcionamento regulatório tem sido reportado como responsável por diversas moléstias como: deformações na fase de desenvolvimento embrionário, manifestação de cânceres de mama, próstata, pâncreas, leucemia, linfoma e cólon. A CSK apresenta uma alta homologia sequencial com as SFKs, mas com disposições distintas de seus domínios. Outro diferencial é a ausência de uma porção terminal contendo uma tirosina que possa ser fosforilada, elemento fundamental para os demais membros das SFKs que podem se autofosforilar. Curiosamente, sua estrutura cristalográfica apresenta conformações distintas classificadas como formas ativa e inativa sendo o objetivo principal desse trabalho, modelar as transições e avaliar os seus possíveis mecanismos fazendo um link de como mudanças conformacionais entre essas duas conformações estão relacionadas com a ação da proteína quinase Csk baseando-se em experimentos *in vitro*. Complementarmente, a proteína adaptadora Grb2 relacionada com regulações de proteínas pertencente ao universo das quinases como a FGFR2, também é alvo do estudo. O conjunto quinase mais proteína adaptadora desempenham um importante papel na transmissão de sinal em cascatas de sinalizações envolvidas em processos celulares fundamentais como proliferação celular, diferenciação e apoptose. Dessa maneira, Grb2 é um alvo em potencial para o estudo de interações com pequenas moléculas de interesse biológico por seus efeitos anticarcinogênicos como a Cumarina 1,2-Benzopirona e Morina (2',3,4',5,7-pentahidroxiflavona). Experimentos *in vitro* comprovam a interação entre a proteína adaptadora Grb2 e o ligante Cumarina e Morina, sendo neste trabalho avaliado a determinação espacial e o conjunto de interações entre os complexos formados.

Palavras-chave: Quinases. C-terminal Src. Csk. Modelos baseados em estruturas. Mudanças conformacionais. Growth Factor Receptor Bound Two. Grb2. Cumarina. Morina. SFKs.

Abstract

The Carboxyl-terminal Src Kinase (Csk) is a protein that plays an essential role for the down-regulation of Src kinase family proteins (SFks). When Csk presents malfunction on its regulation action, several diseases can occur, like deformations during the fetal development, breast, prostate, pancreas, colon cancers, as well leukemia and lymphomas. Csk presents a high sequence homology with other SFK members, but without a similar arrangement of domains. Yet, the Csk lacks a tyrosine tail, so it cannot auto-phosphorylate, but it can phosphorylate their family members. Surprisingly, the crystallography structure of Csk presents two distinct conformations classified as active and inactive. The main objective of this work is to construct a model to evaluate transitions and possible mechanisms involving Csk, making a link between conformational changes and its action. On the other hand, the adaptor protein Growth factor Receptor-Bound protein 2 (Grb2) is related to the universe of protein regulators linked to the protein kinase as the Fibroblast Growth Factor Receptor 2 (FGFR2), being an important target for study. The set kinase and adaptor proteins is important for signaling processes involving cell proliferation, differentiation and apoptosis. Thus, Grb2 is an potential aim for the evaluation of interactions with small anticarcinogen molecules as 1,2-benzopyrone Coumarin and 2',3,4',5,7-pentahydroxyflavon Morin. *In vitro* essays verified the interaction between the Grb2-coumarin and Grb2-morin complexes. Thus, the second part of this work investigate the spatial location for the docking of the Coumarin and Morin into Grb2 and evaluate their interactions.

Keywords: *Kinases. C-terminal Src. Csk. Structure Based Model. Conformational Changes. Growth Factor Receptor Bound Protein 2. Grb2. Coumarin. Morin. SFks.*

Lista de Figuras

Figura 1 - Organização e Mecanismo de regulação das SFKs..	4
Figura 2 - Estruturas cristalográficas da proteína Csk	6
Figura 3 - Perfil de energia livre em função de uma coordenada de reação Q.....	12
Figura 4 - Esboço bidimensional da energia livre F(Q) em função de Q	14
Figura 5 - Mapa de Contatos Shadow.....	17
Figura 6 - Avaliação de transições sem um potencial equiprovável	21
Figura 7 - Identificação de contatos exclusivos de uma conformação.	23
Figura 8 - Avaliação energética das mudanças conformacionais.....	24
Figura 9 - Mecanismos envolvendo a Hélice- α C	26
Figura 10. Alterações energéticas em regiões alvo	28
Figura 11. Deleção dos domínios SH3 e SH2	30
Figura 12 - Deleções de contatos interdomínios.....	33
Figura 13 - Modelagem de um possível ligante em Csk	35
Figura 14 - Representação estrutural da proteína receptora tirosina quinase	41
Figura 15 - Proteína Grb2 em sua forma dimérica	43
Figura 16 - Representação do mecanismo de regulação da proteína quinase FGFR2 através da proteína adaptadora Grb2.....	44
Figura 17 - Classificação das moléculas de Cumarina	46
Figura 18 – Classes e estruturas gerais dos flavonoides e flavonoide Morina	48
Figura 19 - Experimento de STD	51
Figura 20 - Espectro de emissão de fluorescência da proteína Grb2.....	53
Figura 21 – Perfil termodinâmico obtido por fluorescência	55
Figura 22 - SPHGEN, como calcular esferas	58
Figura 23 - Algoritmo Anchor and Grow	61

Figura 24 - Análise estrutural da proteína GRB2 em solução por dinâmica molecular	64
Figura 25 - Clusterização das conformações estruturais de Grb2	65
Figura 26 - Possíveis cavidades para o docking do ligante.	67
Figura 27 – Melhores conformações do Docking da molécula de Cumarina no bolsão contido na interface dimérica	69
Figura 28 – Melhores conformações do Docking da molécula de Morina no bolsão contido na interface dimérica	70
Figura 29 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH2.....	72
Figura 30 – Melhores conformações do Docking da molécula de Morina no bolsão do domínio SH2.....	73
Figura 31 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH3 N-terminal.....	75
Figura 32 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH3 N-terminal.....	76
Figura 33 - Dinâmica molecular da interação dos ligantes com o domínio SH2	78
Figura 34 – Mudanças do domínio SH2 com a presença da Cumarina	79
Figura 35 – Mudanças do domínio SH2 com a presença da Morina.....	80
Figura 36 – Conformações de menor energia retirada da dinâmica molecular.	82
Figura 37 - Interações entre Cumarina e resíduos do domínio SH2 de Grb2.....	85
Figura 38 - Interações entre Morina e resíduos do domínio SH2 de Grb2	86

Lista de abreviações

Å	Angstrom
Csk	C-terminal Src Kinase
FGFR2	Fibroblast growth factor receptor 2
Grb2	Growth factor receptor-bound protein 2
MD	Dinâmica Molecular.
NRTK	Proteína Tirosina Quinase Não Receptora
PDB	Protein Data Bank
PTK	Proteína Tirosina Quinase
Q	Coordenada de reação.
RMN	Ressonância Magnética Nuclear.
RMSD	Desvio Quadrático Médio.
RMSF	Flutuação Quadrática média.
RTK	Proteína Tirosina Quinase Receptora
SH2	Src Homology 2
SH3	Src Homology 3

Sumário

Mudanças Conformacionais da proteína Csk	1
1.1 Introdução.....	2
1.2 Motivação e Objetivos	9
1.3 Fundamentação Teórica.....	10
1.3.1 Enovelamento de proteínas	10
1.3.2 Mudanças conformacionais em proteínas.....	13
1.4 Modelos e métodos	15
1.4.1 Mudanças conformacionais em proteínas.....	15
1.4.2 Coordenadas de reação.....	18
1.4.3 Detalhes das Simulações	19
1.5 Resultados e discussões	20
1.5.1 Construção de um Potencial Equiprovável.	20
1.5.2 Explorando a região da hélice α C.	25
1.5.3 Deleções de domínios.	29
1.5.4 Deleções de contatos interdomínios.	31
1.5.5 Modelagem de um ligante.....	34
1.6 Conclusões	36

Interação entre a proteína Grb2 e as Moléculas Cumarina e Morina	38
2.1 Introdução.....	39
2.1.1 Sinalização celular & proteína FGFR2	40
2.1.2 Growth Factor Receptor-bound 2 - GRB2.	42
2.1.3 A molécula de Cumarina.....	45
2.1.4 A molécula de Morina	47

2.2	Motivações e Objetivos.....	49
2.3	Fundamentação Teórica.....	50
2.3.1	Resultados Experimentais	50
2.3.2	Docking molecular.....	56
2.4	Modelos e métodos	57
2.4.1	Preparação da macromolécula	57
2.4.2	Docking Molecular (UCSF DOCK 6)	57
2.4.3	Dinâmica Molecular do complexo Grb2 e ligantes.....	61
2.5	Resultados e discussões.	63
2.5.1	Identificação de cavidades nos domínios para o docking molecular	66
2.5.2	Dinâmica Molecular.	77
2.5.4	Avaliação da interação melhor ranqueada.....	82
2.6	Conclusões	87
3.	Referências	88
Apêndice A.....		97
Apêndice B.....		99

Capítulo 1

Mudanças Conformacionais da proteína Csk

1.1 Introdução

Grande parte do funcionamento dos processos biológicos em organismos vivos depende de macromoléculas conhecidas como proteínas. Essas macromoléculas são consideradas as mais abundantes, sendo encontradas em todas as células de um organismo. Dessa forma, as proteínas são consideradas vitais para os seres vivos em geral, devido às diversas funções desempenhadas (Nelson & Cox, 2014).

Comumente formadas por 20 diferentes tipos de aminoácidos, com exceções de alguns organismos que podem apresentar 22 diferentes tipos de aminoácidos (Cavalcanti & Landweber, 2003), combinados e ligados covalentemente por ligações peptídicas, as proteínas são divididas em quatro diferentes níveis estruturais. A estrutura primária é composta por uma sequência de aminoácidos formadores da cadeia polipeptídica. As estruturas secundárias são estruturas regulares como α -hélices e folhas β a partir de dobras e padrões específicos na cadeia polipeptídica. A estrutura terciária consiste no arranjo espacial de todos os átomos da proteína, levando em consideração também a geometria de cada estrutura secundária formada. E ainda, quando a proteína possui mais de uma unidade polipeptídica, elas se arranjam em uma geometria específica formando a sua estrutura quaternária. (Branden & Tooze, 1999; Nelson & Cox, 2014).

A organização de cada um de seus níveis estruturais proporciona à proteína diferentes graus de flexibilidade dependentes do seu tamanho e tipo de ligações, podendo ocorrer mudanças conformacionais acessando diferentes estados quando em solução. Essas diferentes organizações de estrutura estão intimamente relacionadas com a sua função biológica, como por exemplo em motores biológicos, nos quais a energia liberada pela quebra da adenosina trifosfato (ATP) é convertida em trabalho mecânico, realizando movimentos e dando origem a sua função biológica (Ronconni, 2011).

Para interagir com uma molécula qualquer, em geral a proteína deve ser capaz de mudar a sua conformação. Tal mudança pode ser mínima, envolvendo o movimento de átomos de um ou mais resíduos, ou brusca, envolvendo o movimento de domínios inteiros de uma estrutura multimérica. Um exemplo de estruturas proteicas com alta

flexibilidade, com suas funções associadas com suas diversas conformações estruturais são as proteínas tirosinas quinases.

As proteínas tirosina quinases (PTKs) são um grupo de enzimas que catalisam a fosforilação de outras proteínas por meio da transferência de um grupo fosfato de ATP ou, em alguns casos excepcionais, de GTP, para resíduos de tirosina. A fosforilação destes resíduos ocorre devido a estímulos extracelulares ou intracelulares, que fornecem um mecanismo altamente eficiente para o controle da atividade de proteínas. (Hubbard & Till, 2000)

As PTKs podem ser divididas em duas famílias: as proteínas tirosinas quinases receptoras (RTKs, do inglês *Receptor Tyrosine Kinases*) como por exemplo, os receptores de crescimento derivados de plaqueta PDGFR (*platelet-derived growth factor receptor*) e a família de fatores de crescimento do fibroblasto FGF (*fibroblast growth factor*), e proteínas tirosinas quinases não receptoras (NRTKs, do inglês *Non-Receptor Tyrosine Kinases*), como as proteínas Src, Jak, Abl, Fak, Fps, Syk, Btk e Csk. (Hubbard & Till, 2000)

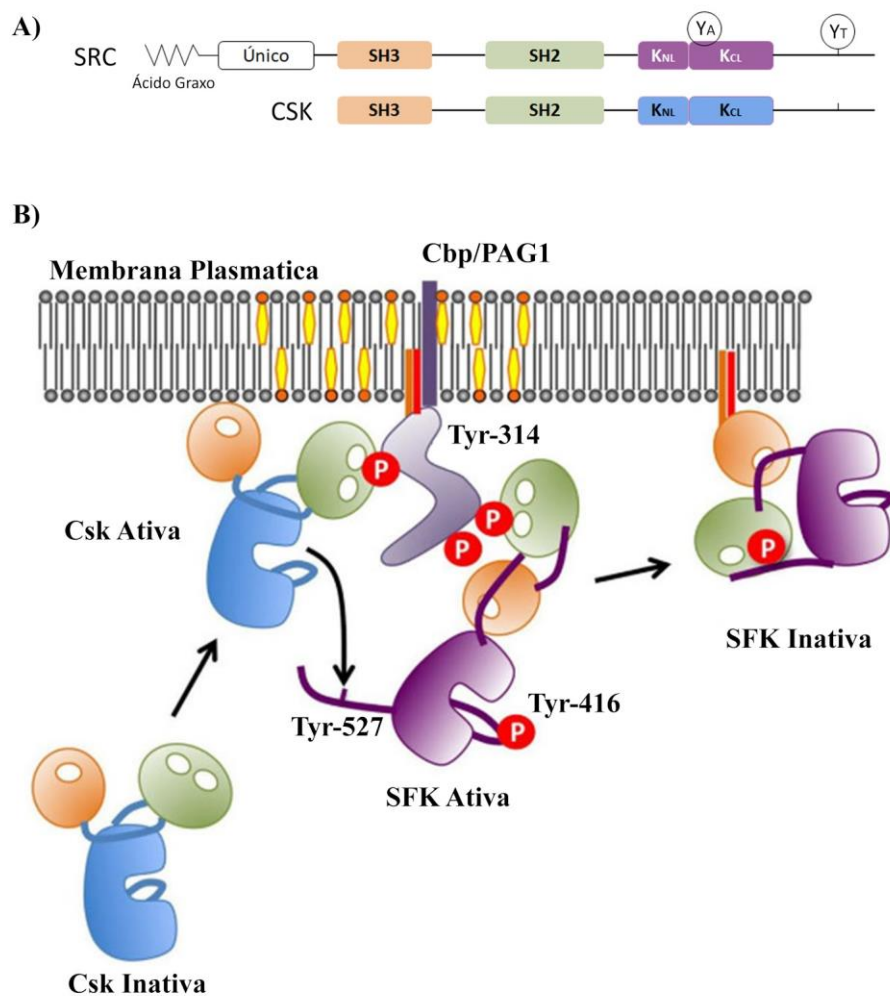
A primeira PTK descoberta foi a Src, uma das proteínas que constitui a subfamília de quinases “*Src Family Kinases*” (SFKs), que em geral atuam no controle do crescimento e diferenciação celular (Chong et al., 2006). SFKs é um grupo composto por nove enzimas, Src, Fyn, Yes, Lck, Hck, Fgr, Lyn, Blk, e Yrk que regulam uma ampla gama de eventos celulares, incluindo crescimento celular, divisão, diferenciação, sobrevivência e apoptose, além de funções especializadas, como respostas imunes, adesão celular, movimento celular e endocitose (Hunter & Cooper, 1985; Thomas & Brugge, 1997).

A família das SFKs contém um ou mais sítios de acilação amino-terminal, necessários para a localização na membrana, um domínio homólogo Src 3 (SH3), que reconhece ligantes ricos em prolina, um domínio homólogo Src 2 (SH2), que reconhece ligantes ricos em fosfotirosina, e um domínio catalítico propriamente dito como domínio quinase subdividido em N-lóbulo e C-lóbulo (Howell & Cooper, 1994; Koegl, Robert, Bergman, Alitalot, & Courtneidge, 1994)

O domínio quinase das SFKs contém dois resíduos críticos para sua regulação: a Tirosina Y416, encontrada no sítio ativo e que sofre autofosforilação para a ativação da quinase, e o resíduo de Tirosina Y527, encontrado em sua cauda C-terminal, que quando fosforilado por outras PTKs, causa rearranjos conformacionais, fazendo com que as SFKs

passem de uma conformação estendida e ativa, para uma conformação fechada e inativa, ilustrado na figura 1.

Figura 1 - Organização e Mecanismo de regulação das SFKs. (A) C-Src apresenta uma região única contendo o sítio de acilação amino-terminal, necessário para a localização na membrana e os resíduos de fosforilação no domínio quinase e cauda C-terminal que modulam sua atividade, sendo esses Tyr-416 (YA) e Tyr-527 (YT) respectivamente. (B) A proteína de membrana Cbp/PAG1 recebe sinais extracelulares ocasionando fosforilação da sua Tyr-314 por proteínas da família SFKs e induz o recrutamento da proteína Csk para membrana. Essa interação ocorre através do domínio SH2 de Csk com a região da Tirosina 314 já fosforilada em Cbp. Csk após recrutada fosforila a Tirosina 527 da cauda C-terminal das SFKs, causando mudanças conformacionais, passando de uma conformação aberta e ativa para sua conformação fechada e inativa.



Fonte: adaptado de (Masato Okada, 2012).

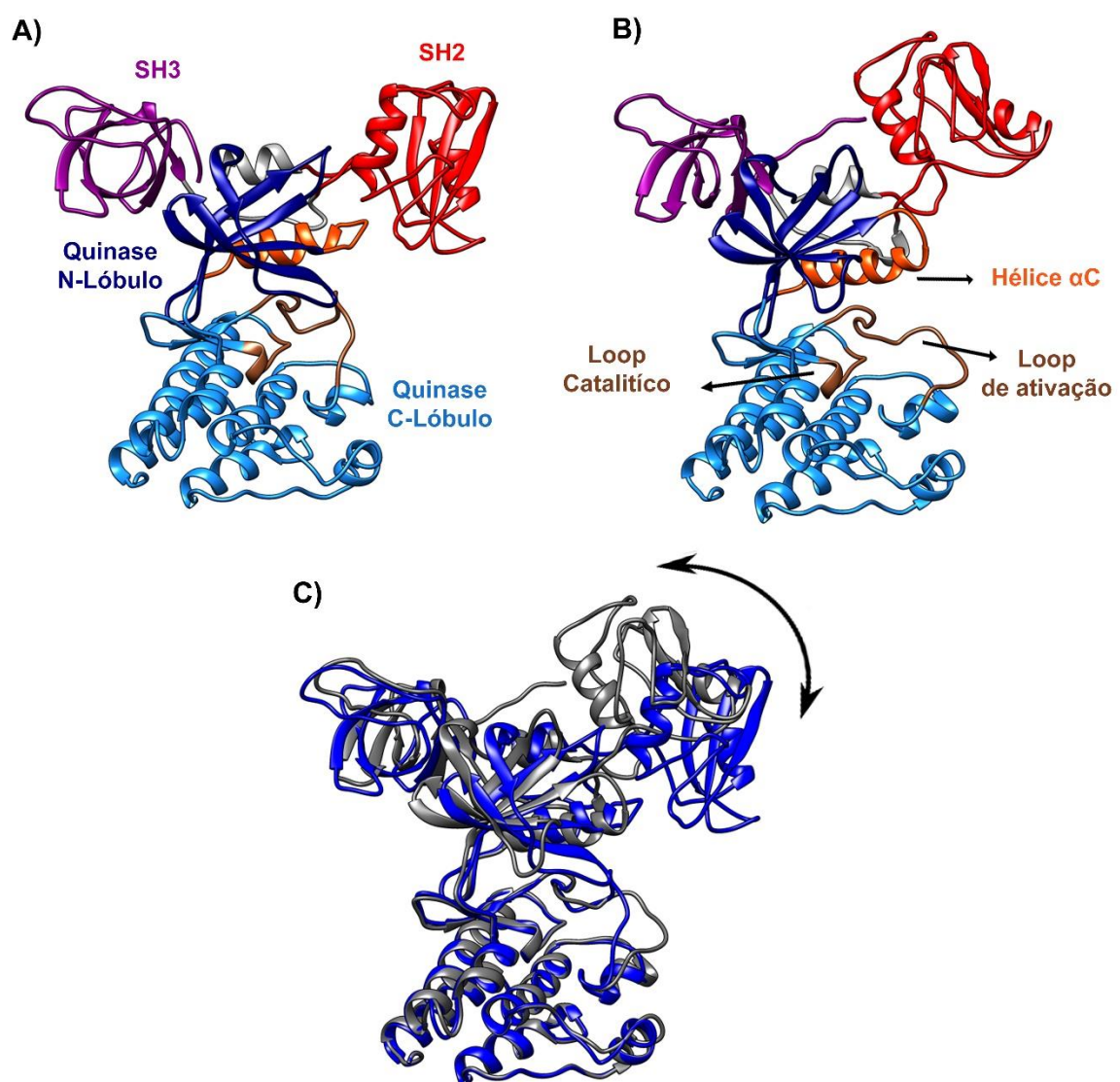
SFKs estão usualmente presentes em seu estado inativo, com sua tirosina Y527 fosforilada interagindo com o domínio SH2. Essa fosforilação é rigorosamente controlada por outra PTK, a Carboxyl-terminal Src Kinase, popularmente conhecida como Csk (Okada et al., 1991; Nada et al., 1993). Csk é uma PTK citoplasmática constituída por um domínio SH3, um SH2 e um domínio Quinase dividido em N-Lóbulo e C-Lóbulo.

Suas estruturas cristalográficas resolvidas por Ogawa e colaboradores (Sakai et al., 2002) consiste em 6 cadeias contidas em uma única célula unitária, e que podem ser divididas em 2 conformações distintas, apresentadas na Figura 2. Em uma das conformações o domínio SH2 de Csk se encontra próximo ao domínio quinase, formando uma configuração mais compacta e sugerida como sua forma ativa. Na segunda conformação, o domínio SH2 é rotacionado resultando em uma conformação estendida, sendo essa sugerida como sua forma inativa. Além do domínio SH2, a primeira hélice do domínio quinase (Hélice α C), colorida em laranja na figura 2, tem uma pequena deformação em sua primeira volta de hélice, evidenciando uma mudança conformacional entre as estruturas cristalográficas.

A perda de função de Csk implica em uma ativação constitutiva das SFKs, acompanhada de moléstias como defeitos no desenvolvimento embrionário e manifestação de cânceres de mama, próstata, pâncreas, leucemia, linfoma e cólon (Nada et al., 1993).

Apesar da identidade estrutural com as SFKs, como mostrado na figura 1A, Csk carece de sítios de acilação amino-terminal, de um sítio de autofosforilação e de uma tirosina reguladora na região C-terminal, características essenciais para a função das SFKs e que conferem à Csk um arranjo estrutural distinto das demais SFKs. À frente de todas essas diferenças, Csk é capaz de fosforilar a Tirosina-527 presente na região C-terminal dos membros da família das SFKs, todavia não sofre autofosforilação (Chong et al., 2006). Portanto, o seu mecanismo regulador difere das SFKs e vem sendo alvo de diferentes estudos.

Figura 2 - Estruturas cristalográficas da proteína Csk. Segundo Ogawa e colaboradores (Sakai et al., 2002), as duas conformações da proteína Csk resolvidas cristalograficamente estão intimamente ligadas com suas funções, sendo sua forma mais compacta a molécula ativa e que realiza suas funções biológicas, e inativa sem funções biológicas quando na sua forma estendida. Suas estruturas compacta e estendida estão representadas em (A) e (B) respectivamente, coloridas por seus domínios, Loop Catalítico, Loop de ativação e com destaque para a região de estrutura secundária Hélice α C, evidenciando que as mudanças conformacionais não ocorrem somente com o movimento do domínio SH2. Em (C) é representado a sobreposição das cadeias B e C, coloridas em azul e cinza respectivamente.



Fonte: Figura elaborada pelo autor.
PDBid: 1K9A, cadeias B e C.

Estudos envolvendo o domínio SH2 de Csk mostram a relevância deste domínio para a atividade de Csk; interações entre fosfotirosinas de algumas moléculas com este domínio podem aumentar a catálise de Csk, como por exemplo, a proteína de ligação de Csk, Cbp/PAG (no inglês, *Csk Binding protein*) (Tanaka et al., 2013), a qual recruta Csk para a membrana em regiões estratégicas, possibilitando essa a atuar diretamente sobre as quinases da família Src (Davidson, Bakinowski, Thomas, Horejsi, & Veillette, 2003). Avaliações detalhadas de mutagênese mostram também que a quebra de suas interações com o N-lóbulo, reduz significativamente a atividade catalítica de Csk, implicando em regulações descontroladas em SFKs (Huang, Wang, Brown, & Sun, 2009; Sondhi & Cole, 1999; Sun & Budde, 1999).

Ainda relacionado com o domínio SH2, estudos mostraram que a sua deleção reduz a atividade catalítica de Csk (isto é, k_{cat}/K_m) por 30 vezes (X. Lin, Ayrapetov, Lee, Parang, & Sun, 2005). No domínio SH2 da Csk, encontra-se uma ligação dissulfeto que não está presente em outros domínios SH2 conhecidos. Estudos de RMN mostram que a ligação dissulfeto acessível ao solvente no domínio SH2 isolado é altamente estável e longe do pequeno lobo do domínio da quinase. Entretanto, a quebra dessa ligação produz efeitos em outras regiões da proteína, em especial no lóbulo N-terminal do domínio Quinase onde encontra-se a hélice αC , afetando a função catalítica de Csk evidenciando novamente a importância da comunicação interdomínios e destacando o envolvimento da hélice αC com a atividade proteica (Mills et al., 2007).

O domínio SH3 também apresenta relação com a atividade da Csk, sendo que sua remoção diminuiu sua atividade em 11 vezes (X. Lin, Ayrapetov, Lee, Parang, & Sun, 2005). Também, mutações nos resíduos E13A e D34A pertencentes ao domínio SH3 e que fazem interações interdomínios, resultando na redução da atividade de Csk, evidenciando a importância da comunicação entre os diferentes domínios da proteína (Barkho et al., 2015).

É possível também associar a atividade da proteína Csk com sua interação com outras moléculas devido a sua identidade estrutural com as SFKs, como é o caso da *Staurosporine* (Lamers, Antson, Hubbard, Scott, & Williams, 1999; Meggio et al., 1995). Essa molécula interage com algumas das proteínas da família das SFKs próxima à região da hélice αC , afetando sua atividade. A molécula *Staurosporine* (antibiótico AM-2282 ou STS) é um produto natural originalmente isolado em 1977 da bactéria *Streptomyces*

Staurosporeus (Omura et al., 1977). Sua principal atividade biológica é a inibição das proteínas quinases impossibilitando a ligação de um ATP. Isto é possível pois há uma forte afinidade entre a *Staurosporine* e Quinase no local de ligação do ATP (Karaman et al., 2008). A análise estrutural das interações entre diferentes quinases e a molécula de *Staurosporine* mostram que o bolsão de ligação é preservado devido aos átomos da cadeia principal conservados em suas posições relativas, contribuindo para a promiscuidade da molécula (Boggon et al., 2012; Kinoshita, Matsubara, Ishiguro, Okita, & Tada, 2006; Schiering et al., 2003).

Estudos recentes apresentados por Jamros e colaboradores (Jamros et al., 2010, 2012) indicam que em solução existe um equilíbrio entre conformações estendidas e colapsadas da Csk, sugerindo que as existências desses *ensembles* sejam fundamentais para sua dinâmica de ação. Ainda, na presença dos nucleotídeos AMP-PNP (um análogo do ATP não hidrolisável) ou ADP existe um deslocamento desse equilíbrio, sempre exigindo a combinação de 2 ou mais estruturas representativas para a reconstrução adequada dos dados de espalhamento de raios-X à baixo ângulo. Complementarmente, foi reportado a importância dessas moléculas no processo de regulação reversa, permitindo a fosforilação da proteína de ligação Cbp/PAG ou a c-Src pela Csk (Jamros et al., 2012). A avaliação dos efeitos causados por estas diferentes regiões da Csk sugere que os estudos dos mecanismos da transição são muito importantes para o entendimento da função da Csk.

Essas descobertas sobre o comportamento cinético da Csk em solução sugerem uma dinâmica cooperativa entre os domínios assim como a relação das transições pelos diferentes *ensembles* possíveis com a sua capacidade de ativar o domínio C-terminal de tirosina quinase. Portanto, identificar regiões responsáveis por facilitar ou limitar as transições indicariam um ponto de partida para o desenho racional de novas drogas que atuam para manter a ação desejável da Csk no organismo, oferecendo possíveis soluções de tratamento para as doenças já mencionadas.

1.2 Motivação e Objetivos

A proteína Carboxyl-terminal Src Kinase, Csk, é uma proteína tirosina quinase com função de reguladora essencial das proteínas quinases da família das SFKs. A Csk atua fosforilando um resíduo de tirosina (Tyr-Y527) presente na região C-terminal de SFKs, induzindo mudanças conformacionais e inibindo sua atividade proteica. Apesar de sua homologia estrutural com as SFKs, a Csk carece de regiões e resíduos importantes para a regulação de SFKs, diferindo quanto ao seu mecanismo regulador e que vem sendo alvo de diversos estudos.

Os resultados cristalográficos mostram diferentes conformações estruturais nas células unitárias, apontando para o pressuposto de que haviam dois estados, um ativo e outro inativo. Contudo, diversos trabalhos experimentais, por exemplo empregando Ressonância Magnética Nuclear (RMN), deleção de regiões ou domínios, mutação de resíduos de aminoácidos e interação com outras moléculas evidenciam a comunicação interdomínios e a extrema importância da cinética da Csk, indicando uma relação com sua função. Aqui questiona-se a hipótese da atividade da Csk depender de todo seu mecanismo de transição, ao invés da tradicional abordagem relativa a dois estados exclusivamente, sendo este o objetivo principal.

Os desafios secundários ao objetivo principal, foi explorar regiões que restringem a livre ocorrência de transições, como os mecanismos físicos de *cracking* e *back-tracking*, por exemplo. Essas análises possibilitam o entendimento da intercomunicação entre os domínios e da atuação de diferentes ligantes nos efeitos alostéricos já conhecidos para Csk. Além disso, estas análises são um ponto de partida para a identificação, desenvolvimento e modelagem de compostos químicos que venham interagir em regiões específicas identificadas de modo desejável, controlando assim a regulação negativa da Csk sobre os outros membros das SFKs.

1.3 Fundamentação Teórica

1.3.1 Enovelamento de proteínas

Na década de 60, Anfinsen por meio de experimentos físico-químicos, sugeriu que conhecendo a sequência de aminoácidos, seria possível determinar o estado funcional de uma proteína (estado nativo), sendo esse, o mínimo global da energia livre (Anfinsen, 1973; Anfinsen, Haber, Sela, & White, 1961). O experimento realizado por Anfinsen consiste na desnaturação da Ribonuclease Pancreática Bovina A modificando as condições fisiológicas pela adição de ureia ao sistema e controlando a sua desnaturação por meio da contagem de 4 pontes dissulfeto presentes na estrutura nativa da proteína. O resultado mostrou que a proteína depois de desenovelada, desde que colocadas em sua condição fisiológica novamente, espontaneamente voltava a se enovelar. Anfinsen estabelece assim sua “Hipótese Termodinâmica” afirmando que as proteínas em seu estado nativo e condições fisiológicas normais, está em seu estado de mínima Energia Livre, e que essa conformação é, portanto, estabelecida apenas por interações Inter atômicas (Anfinsen, 1973). Isso significa que toda a informação para a proteína alcançar seu estado funcional está codificada na sequência de aminoácidos, não requerendo outros sistemas biológicos acessórios para o enovelamento.

Em 1968, Cyrus Levinthal questionou sobre o tempo necessário para uma proteína alcançar a sua estrutura nativa e funcional, considerando que ela tivesse que acessar todas as suas conformações possíveis. O problema pode ser exemplificado com a seguinte suposição: se o processo de enovelamento de proteínas percorre caminhos aleatórios, e se pelo menos um Picosegundo fosse gasto para acessar uma conformação, o tempo necessário para o enovelamento de uma proteína com cem aminoácidos e duas conformações (caso mais simples), para cada ligação peptídica, seria da ordem de 10^{18} segundos, tempo comparável com a idade do universo, o que ficou conhecido como o Paradoxo de Levinthal. Levinthal concluiu então que para que o enovelamento de uma proteína ocorresse em tempo hábil seria necessário um caminho determinado, definido por um “*pathway*” (Zwanzig, Szabo, & Bagchi, 1992).

Passados quase 50 anos de trabalho e com a associação de diversas teorias como a de *spin-glass* e de superfície de energia (do inglês, *Energy Landscape*) (Chan & Dill,

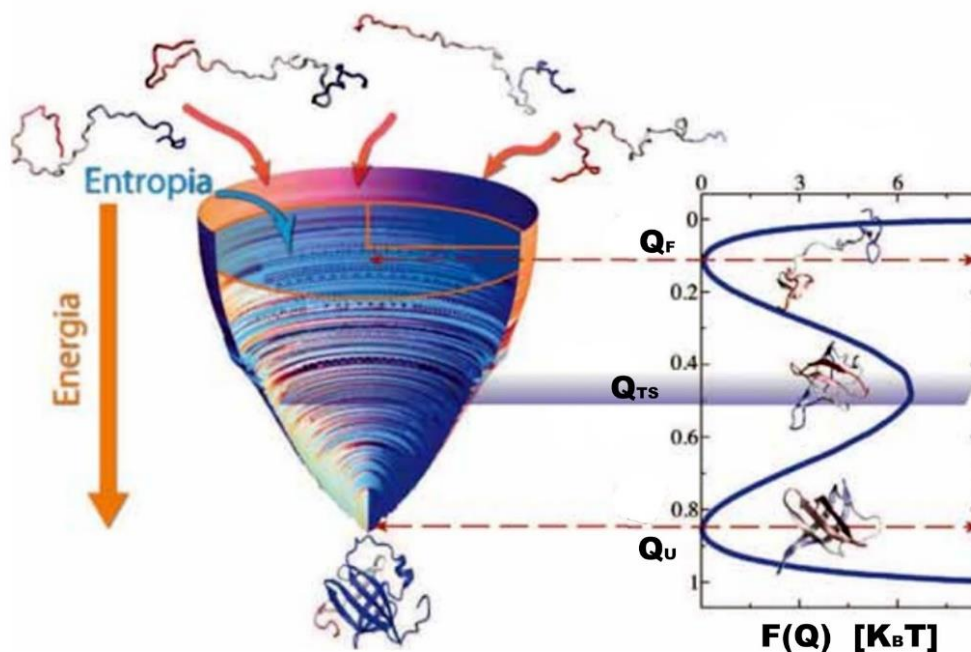
1989), foi possível entender o que conhecemos hoje como “Teoria de enovelamento de proteínas” e “Funil de enovelamento”. Esses avanços foram desenvolvidos durante a década de 80 por Peter Leopold et Al. (Chan & Dill, 1989; Taketomi, Ueda, & Gō, 1975) que introduziu o conceito de funil de energia como consequência dos avanços de outros pesquisadores (Dill & Chan, 1997; Karplus, Šali, & Shakhnovich, 1995). Nessa teoria, entende-se a existência de um balanço entre entropia configuracional e energia do sistema, que direciona a proteína para o seu mínimo de energia livre. Assim, o processo de enovelamento da proteína pode assumir diferentes rotas, sendo um tratamento estatístico baseado nos princípios da termodinâmica e física estatística (Dill & Chan, 1997).

Conceitualmente, a superfície de energia é representada por uma forma afunilada como mostrado na Figura 3, onde encontra-se os valores de energia no eixo das ordenadas com valores crescentes para cima, e valores de entropia no eixo das abscissas, com valores crescentes para as laterais, tendo, portanto, em seu topo o *ensemble* da proteína desenovelada e no fundo do funil o *ensemble* da conformação enovelada. A proteína é então direcionada para seu estado nativo através de rotas que minimizam a energia livre.

Nessa superfície de energia podem existir mínimos locais de energia representados por pequenos poços no funil, os quais podem “armadilhar” a proteína ao longo do processo de enovelamento. Essas armadilhas são conhecidas como a rugosidade do sistema e estão relacionadas com o tempo de enovelamento e acessibilidade da proteína ao estado nativo, sendo importante para essa avaliação a relação T_f/T_g (temperatura de enovelamento dividido por temperatura de *glass* ou do estado vítrio). Ressaltando que o comportamento cinético e termodinâmico do enovelamento é influenciado pelo relevo da superfície de energia. Nos casos em que as superfícies são irregulares (ou rugosas) há uma grande competição entre as interações favoráveis e desfavoráveis da proteína, sendo as interações desfavoráveis chamadas de “frustração” (J. D. Bryngelson, Onuchic, Socci, & Wolynes, 1995; Leopold, Montal, & Onuchic, 1992). De acordo com o princípio da mínima frustração, introduzido por Bryngelson e Wolynes no estudo de heteropolímeros aleatórios (J. D. D. Bryngelson & Wolynes, 1987; J. D. Bryngelson et al., 1995), proteínas são sistemas otimizados, sendo

que uma interação é dita frustrada quando ela não é localmente satisfeita em termos de sua energia.

Figura 3 - Perfil de energia livre em função de uma coordenada de reação Q . O funil de energia tem em seu topo estruturas de proteínas desenoveladas, estabelecendo essa região como sendo de alta entropia e alta energia. Sua forma afunilada representa as diferentes rotas de energia para a estrutura nativa, sendo o “bico” do funil uma região de mínima energia. Os parâmetros apresentados na imagem são: Q_U = Conformação desenovelada. Q_{TS} = Conformação no estado de transição, Q_F = Conformação enovelada e $F(Q)$ a energia livre em função da coordenada de reação Q .



Fonte: adaptada de (Wang et al., 2012).

As frustrações de uma proteína podem ser classificadas como energéticas ou topológicas (Clementi, Nymeyer, & Onuchic, 2000). A frustração energética é associada com a sequência de aminoácidos que compõem a proteína e com a energia de interações entre esses aminoácidos, na qual todas as interações energeticamente favoráveis não podem ser satisfeitas (Dill & Chan, 1997). As frustrações geométricas ou topológicas estão relacionadas com o empacotamento ou estruturas secundárias, nas quais por exemplo mudanças conformacionais entre microestados exijam

desenovelamento de pequenas regiões da proteína (Clementi et al., 2000; Nelson & Onuchic, 1998; Onuchic, Socci, Luthey-Schulten, & Wolynes, 1996), processo também conhecido também como *cracking*, ou ainda quebra de contatos favoráveis ao longo do enovelamento, conhecido como *back-tracking*.

Estudos computacionais do processo de enovelamento exigem modelos simplificados porque, quase sempre, demandam um alto custo computacional. Tais modelos geralmente são representações básicas das proteínas envolvendo um número reduzido de parâmetros. Uma linha computacional sugere trabalhos teóricos utilizando modelos minimalistas de rede plana ou cúbicas, evoluindo essas ideias para modelos fora da rede apelidados como modelos Gö (Taketomi et al., 1975). Esses modelos, conhecidos hoje como modelos baseados em estrutura, tem a energia potencial mínima ajustada para o estado nativo da proteína, sendo considerado que resíduos próximos sempre interagem atrativamente, na sua modelagem mais simples.

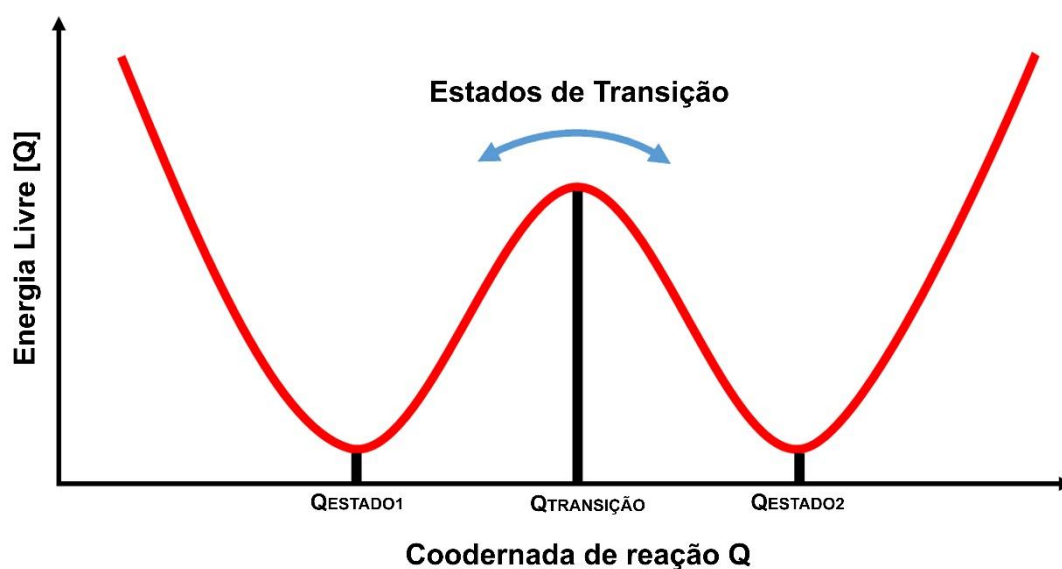
1.3.2 Mudanças conformacionais em proteínas

A teoria de superfície do relevo de energia foi apoiada em hipóteses para elucidar o problema do enovelamento de proteínas, isto é, dizem respeito ao caminho realizado pela proteína para sair de um estado desenovelado e alcançar o seu estado nativo. Essa modelagem de dois estados, todavia consideram que a proteína tem 1 único estado enovelado, por exemplo a proteína em sua conformação cristalográfica. Todavia, existem proteínas que apresentam um regime de transições entre uma ou mais conformações representativas em solução. Nesses casos mudanças conformacionais podem ocorrer, como é o caso da proteína quinase Csk.

Ideias do modelo, com a mesma ideia do funil de energia, podem ser aproveitadas por meio de adaptações que permitam o acesso entre dois ou mais funis, indicando mínimos de energias iguais para estruturas diferentes (Whitford, Miyashita, Levy, & Onuchic, 2007), cada um representando um *ensemble* estável da proteína, como ilustrado na figura 4. A construção desse modelo requer duas estruturas estáveis já conhecidas da mesma proteína. Perturbações graduais são adicionadas na topologia de uma dessas estruturas até se ter o equilíbrio, onde os mínimos de energia entre os dois

ensembles estáveis se igualam, permitindo assim a ocorrência de transições conformacionais durante simulações computacionais.

Figura 4 - Esboço bidimensional da energia livre $F(Q)$ em função de Q . A linha vertical em $Q_{\text{TRANSIÇÃO}}$ marca a região de barreira de energia que corresponde à transição de um dos estados estáveis da proteína (Q_{ESTADO1}) para o outro estado estável (Q_{ESTADO2}). Os mínimos de energia livre foram modelados para o mesmo valor, indicando iguais probabilidades de cada estado ser acessado durante dinâmicas moleculares, proporcionando um equilíbrio dos diferentes estados da proteína.



Fonte: Elaborada pelo Autor.

1.4 Modelos e métodos

1.4.1 Mudanças conformacionais em proteínas

Proteínas são macromoléculas complexas que requerem um tratamento adequado para o entendimento de seu funcionamento e de processos biológicos em que elas estejam envolvidas. Modelagens ricas em detalhes muitas vezes não atendem os recursos disponíveis para a avaliação de aspectos globais devido ao alto custo computacional envolvido. Uma abordagem promissora nessa situação é o emprego de modelos simplificados que capturem propriedades gerais do sistema, possibilitando que simulações computacionais ocorram em um tempo razoável e com resultados capazes de promover o cálculo de propriedades termodinâmicas de interesse. Vale salientar que outros modelos simplificados podem ser utilizados, como a análise de modos normais (modelo de rede elástica), com menor custos computacionais, todavia com perda de informação relevante ao longo do processo.

Modelos baseados em estrutura (*Structure Based Models*, SBM), são construções em que a energia potencial do sistema de interesse é determinada exclusivamente pelas propriedades geométricas da conformação fornecida. Eles apresentam um bom acordo com os resultados experimentais de ϕ -values (Naganathan & Munoz, 2010) e RMN (Mills et al., 2007) para a descrição do processo de enovelamento de proteínas e de transições conformacionais, quando analisados em relação a valores medidos no estado de transição, em relação ao estado enovelado e desenovelado (Clementi, Jennings, & Onuchic, 2001; Lindberg et al., 2002).

Existem várias representações possíveis empregando modelos baseados em estrutura, dependendo da forma como os átomos são selecionados. Aqui, foi empregado um modelo baseado em estrutura em que os resíduos de aminoácidos foram representados apenas pelos seus carbonos α .

A captura de propriedades geométricas foi realizada pela função energia potencial, construída empregando informações contidas na estrutura cristalográfica fornecida, conforme já mencionado. A energia potencial das novas conformações (Γ) foi determinada em relação a sua conformação nativa (Γ_0), sendo dada por:

$$\begin{aligned}
U(\Gamma, \Gamma_0) = & \sum_{covalentes} \frac{1}{2} \epsilon_r (r - r_0)^2 + \sum_{angulares} \frac{1}{2} \epsilon_\theta (\theta - \theta_0)^2 + \\
& + \sum_{diedral} \left\{ \epsilon_\varphi [1 - \cos(\varphi - \varphi_0)] + \left[\frac{1}{2} \epsilon_\varphi [1 - \cos(3(\varphi - \varphi_0))] \right] \right\} \\
& + \sum_{i < j-3} \left\{ \epsilon_{LJ}(i, j) \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] \right\} + \sum_{ij} \epsilon_{LJ-rep}(i, j) \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}
\end{aligned}
\tag{eq. 1}$$

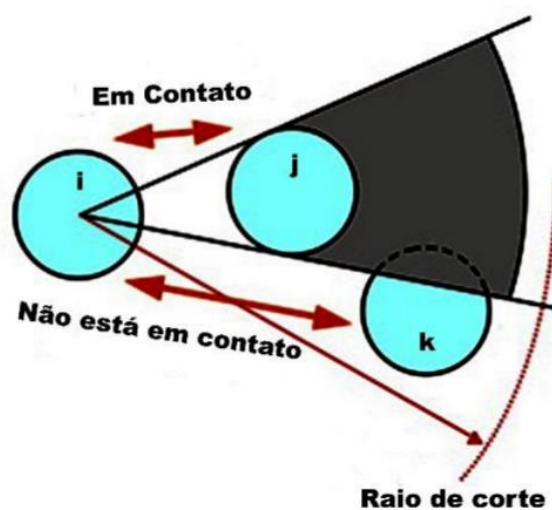
Na expressão, o primeiro termo é um potencial harmônico que representa a ligação entre dois carbonos α adjacentes e r representa a distância entre dois carbonos α ligados covalentemente. O segundo termo da expressão é um potencial harmônico angular formado por três Carbonos α consecutivos, onde θ é o ângulo formado por três resíduos i , $i+1$ e $i+2$ da estrutura nativa. Estes dois primeiros termos modelam propriedades intrínsecas da química orgânica, isto é, são propriedades geométricas que independem da estrutura nativa. O terceiro termo refere-se à torção realizada na cadeia, onde φ_0 é o ângulo diedro formado por quatro carbonos consecutivos. O quarto termo representa as interações (de curto alcance definidas por um mapa de contatos) de Lennard-Jones (potencial 12-10)¹, onde σ_{ij} é o valor da distância entre os carbonos nativos i e j e r_{ij} é a distância entre dois átomos quaisquer. O quinto e último termo contabiliza a repulsão entre os contatos que não estão presentes na estrutura nativa da proteína, sendo equivalente a parte repulsiva do potencial de LJ, exceto por tratar os resíduos que não estão contemplados no mapa de contato. A aproximação máxima entre os carbonos alfa caracteriza o volume ocupado por um aminoácido que é de $\sigma_{NC} = 4 \text{ \AA}$. (Clementi et al., 2000)

¹ O Potencial 10-12 é usado porque as distâncias de contatos nativas σ_{ij} são maiores no modelo C α . Assim esse potencial nesse tipo de modelagem aumenta a especificidade, reproduzindo uma reação entre dois estados, conforme proposto na figura 4.

As constantes ϵ_r , ϵ_θ , ϵ_ϕ , e ϵ_{NC} são os pré-fatores energéticos com valores de 100, 20, 1 e 1 respectivamente, e são dados em unidade de ϵ_C , que é a energia de interação entre 1 par de contatos, apresentando um valor de 1 kcal/mol aproximadamente (Clementi et al., 2000). Estes parâmetros podem ser ajustados com certa flexibilidade (Oliveira, Schug, & Onuchic, 2008) devido a robustez do modelo. Os parâmetros r_0 , θ_0 , ϕ_0 e d_{ij} são os valores obtidos da estrutura nativa.

O valor de d_{ij} foi obtido por meio do cálculo do mapa de contatos, para realizar esse cálculo foi utilizado o servidor SMOG@ctbp (Noel et al., 2016) com o algoritmo *Shadow Contact Map* (SCM) (Noel, Whitford, & Onuchic, 2012), que consiste em uma abordagem de raio de corte com a vantagem de associar um filtro de seleção dos contatos removendo efeitos blindagem. O método pode ser compreendido supondo uma fonte de luz imaginária localizada no centro de um dado átomo i , todos os átomos ao seu redor e dentro do raio de corte que não tenha sombra provenientes de outros átomos a sua frente e que barrem a luz, são considerados contatos, como mostrado na Figura 5.

Figura 5 - Mapa de Contatos Shadow. O átomo de referência considerado como sendo uma fonte de luz está destacado pela letra i . Todos os átomos dentro do raio de corte e que não tem sombra de outros átomos projetados em si, são considerados como contatos. Na representação, o átomo j é definido como contato e o k é definido como não contato.



Fonte: Imagem adaptada de (Noel et al., 2012).

1.4.2 Coordenadas de reação

Coordenada de reação é um parâmetro fenomenológico necessário para a descrição e identificação da mudança conformacional por meio de uma grandeza física relacionada. Se adequadamente escolhida, ela é capaz de descrever a dinâmica dessas mudanças em função de parâmetros físicos do sistema. A coordenada escolhida será responsável pela boa descrição dos processos de interesse envolvidos e dos parâmetros calculados durante as análises. Atualmente existe uma grande discussão entre quais seriam as coordenadas de reação mais adequadas para cada caso no estudo de proteínas, assim foram escolhidos os parâmetros mais comuns da literatura como ponto de partida.

Um dos parâmetros utilizado para a avaliação de mudanças conformacionais na proteína Csk foi a distância entre dois domínios ou resíduos da proteína. Essa coordenada descreve o quão distante está o centro de massa no caso de domínios, ou o quão distante estão os resíduos durante as simulações. Uma vez que o domínio SH2 sofre o deslocamento proposto entre as diferentes estruturas (Figura 2) e o domínio quinase C-terminal não sofre deslocamentos, a distância entre esses dois grupos foi utilizada como parâmetro diferenciando os dois *ensembles* da proteína.

Outro parâmetro utilizado como coordenada de reação para a avaliação de transições na Csk, foi a Raiz do Desvio Quadrático Médio (RMSD). Essa medida refere-se ao quanto as posições atômicas de uma determinada estrutura está se desviando de uma estrutura de referência e é calculada por meio da equação 2.

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[(x_f - x_i)^2 + (y_f - y_i)^2 + (z_f - z_i)^2 \right]} \quad (\text{eq. 2})$$

Onde N é o número total de átomos no sistema, x , y e z refere-se ao sistema de coordenadas espaciais utilizado, e os índices f e i representam as coordenadas finais e as coordenadas de referência (ou iniciais), respectivamente. Vale ressaltar ainda que este parâmetro pode ser utilizado tanto com a proteína completa, quanto com domínios ou regiões específicas, como por exemplo hélices alfa e folhas betas.

Diedros da proteína também podem ser empregados como coordenadas de reação para avaliações locais mostrando o comportamento de regiões específicas e seus mecanismos, indicando se a região é importante ou não para que ocorra mudanças conformacionais.

1.4.3 Detalhes das Simulações

As informações estruturais da proteína Csk foram retiradas do banco de dados *Protein Data Bank* (PDB) (Bernstein et al., 1977) utilizando o código de acesso 1K9A:B para a estrutura compacta e 1K9A:C para estrutura aberta, nas quais encontram-se as informações referentes às coordenadas espaciais de cada átomo. Com a estrutura nativa, foi realizado a montagem do mapa de contato utilizando o servidor SMOG@ctbp (Noel et al., 2016) com o algoritmo *Shadow Contact Map* (Noel et al., 2012), obtendo assim o arquivo de topologia contendo todas as informações de contatos com seus parâmetros energéticos (ângulos e diedros) e de estrutura a ser utilizado durante as simulações.

As simulações foram conduzidas utilizando pacote de programas GROMACS, versão 4.5.7 (Pronk et al., 2013), utilizando um integrador estocástico e o acoplamento térmico de Berendsen (Lemak & Balabaev, 1994). Ao todo foram realizados 1×10^9 passos de simulação com um intervalo de tempo dt de 0.0005 ps totalizando um tempo de simulação de 500ns. As coordenadas espaciais e de energia foram salvas a cada 4000 passos, com um total de 250000 frames. Por se tratar da versão Ca de SBM esse tempo é pelo menos 4 vezes superior ao tempo real (Karplus et al., 1995).

1.5 Resultados e discussões

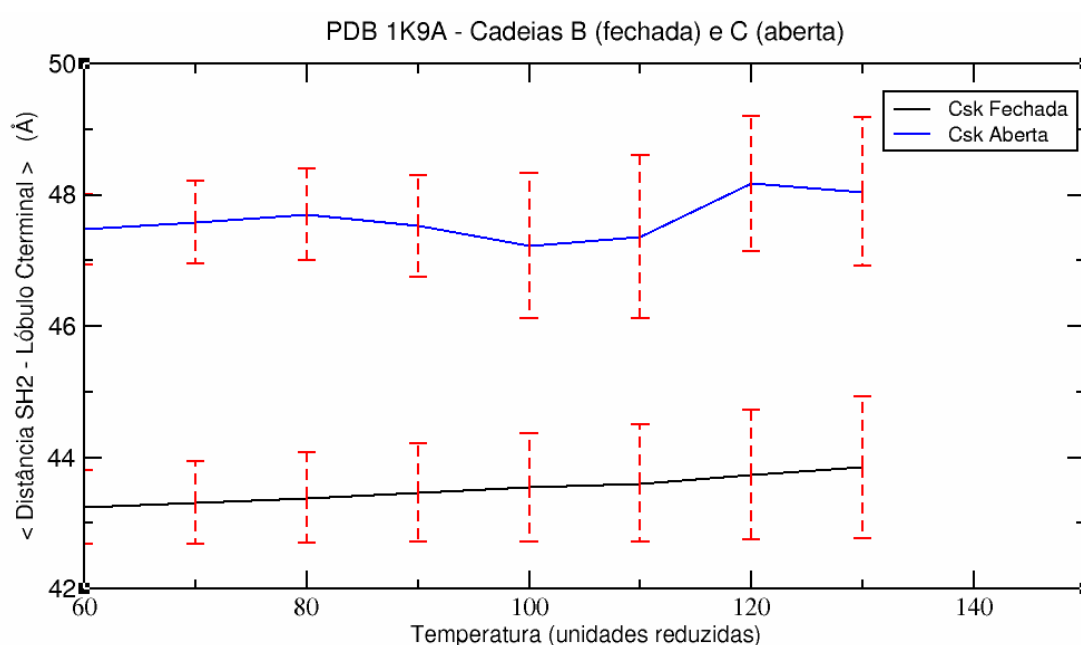
1.5.1 Construção de um Potencial Equiprovável.

Os resultados cristalográficos (PDBid 1K9A) (Sakai et al., 2002) e de SAXS (Jamros et al., 2010) indicam que a proteína Csk apresenta mais de uma conformação nativa possível, especialmente pelos resultados em solução em que é mostrado um equilíbrio entre conformações. Esse equilíbrio é dado como importante para sua função e foi verificado seu deslocamento na presença de AMP-PNP e ADP (Barkho et al., 2015; Jamros et al., 2012).

Quando foram realizadas simulações empregando os modelos baseados em estruturas partindo de qualquer uma das duas diferentes conformações, 1K9A:B (conformação mais compacta) e 1K9A:C (conformação mais aberta), os resultados mostram que as estruturas oscilam em torno das suas conformações iniciais mesmo com o aumento de temperatura. Ou seja, as informações cristalográficas não foram suficientes para que, durante as simulações computacionais, partindo de umas das estruturas, esta alcance a outra, como mostrado na Figura 6.

Essa é uma importante verificação para confirmar que as transições não acontecem naturalmente utilizando quaisquer uma das conformações como ponto de partida, mesmo com o aumento da temperatura. Assim, foram necessárias adequações na energia potencial do sistema para que as transições pudessem ocorrer livremente, o que foi obtido pela inclusão de contatos e ajustes nas intensidades das interações. Esse novo potencial é denominado dual-Gö ou potencial com estados equiprováveis.

Figura 6 - Avaliação de transições sem um potencial equiprovável. Simulações com os modelos baseados em estrutura da proteína Csk partindo das conformações fechada (PDB id: 1K9A:B) e aberta (PDB id: 1K9A:C) da proteína obtida por cristalografia são realizadas em diferentes temperaturas. Nota-se que, utilizando como parâmetro de avaliação a distância entre o domínio SH2 e o domínio Quinase C-terminal, as simulações que partem da estrutura aberta, permanece em torno de sua estrutura inicial e não realizam transições para a outra forma. O mesmo ocorre para as simulações que partem da conformação fechada. Todo o processo foi realizado partindo de uma temperatura de 60 em unidade reduzidas² até a temperatura de 130 em unidades reduzidas pelo fato de que nessa faixa não ocorre o desenovelamento proteico. Essas temperaturas correspondem ao intervalo entre a temperatura de “congelamento” (*glass temperature*) e de fusão (*melting temperature*).



Fonte: Elaborada pelo autor.

Para a construção do potencial dual-Gö, a estrutura da proteína Csk de conformação extensa (cadeia C, PDB 1k9A) foi adotada como estrutura de referência. Os contatos exclusivos da estrutura fechada (cadeia B, PDB 1k9A) foram identificados e as distâncias entre esses átomos foram calculadas utilizando os arquivos de coordenadas das duas diferentes conformações. Os contatos em que as distâncias

² No pacote utilizado para dinâmica molecular, Gromacs, $k_B = 0,00831451$ e não 1. Assim, para ter uma temperatura reduzida de 1, você deve usar uma temperatura de $1 / 0,00831451 = 120,2717$.

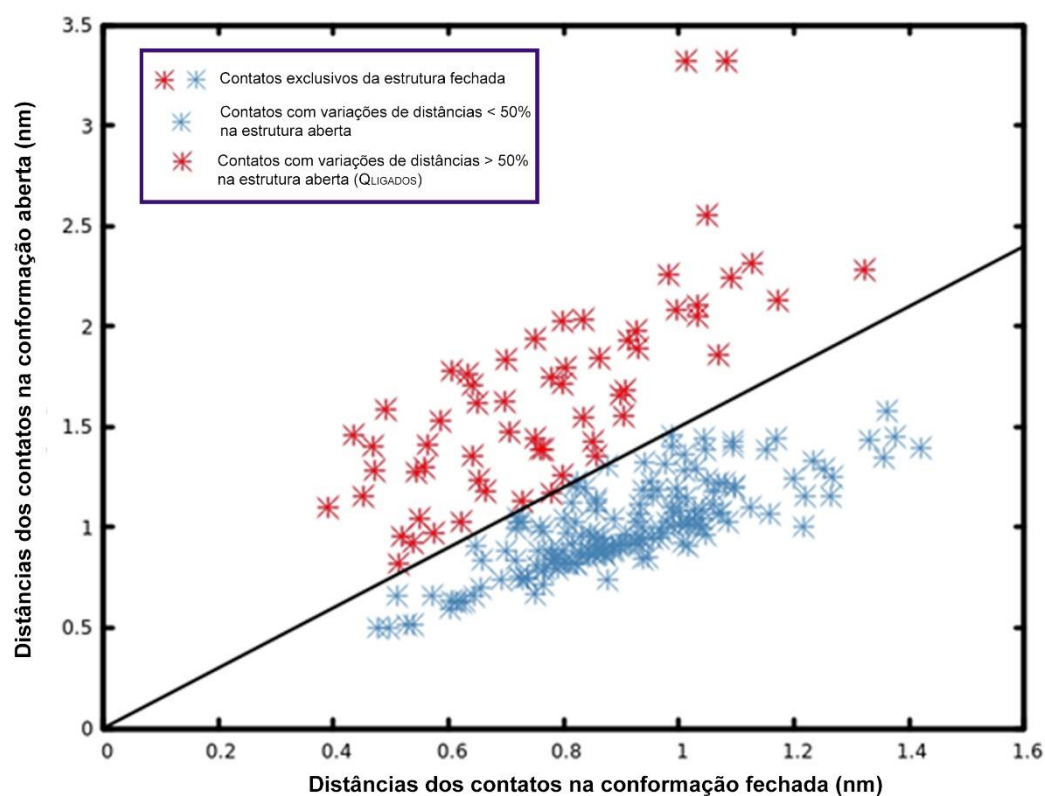
mudaram em 50% ou mais, considerando as duas conformações (Q_{LIG} na figura 7), foram dados como contatos de interesse e prováveis responsáveis pelas transições. Esses contatos foram inseridos no arquivo de topologia da estrutura de referência, constituindo assim um novo arquivo de topologia formado por todos os contatos da conformação aberta mais os contatos de interesse da conformação fechada.

Os novos contatos adicionados ao arquivo de topologia influenciaram na energia total do sistema, exigindo então um ajuste para o modelo. Esses ajustes foram realizados multiplicando os valores do pré-fator de energia de contato (coeficiente ϵ_c da eq 1) por uma constante η igual à 2,1 para todos os contatos que foram adicionados. Para determinar o valor ideal de η como 2,1, foram realizados diversos testes com diversos valores de η e para cada um destes foram realizadas simulações computacionais contemplando diferentes valores de temperaturas. O resultado ideal é aquele em que as duas estruturas do cristal de Csk são acessadas com igual frequência ou tempo de permanência nas populações, dentre as temperaturas de congelamento e desenovelamento.

Para analisar os resultados de cada uma das simulações, foram utilizadas as coordenadas de reações: distância entre o domínio SH2 ao lóbulo C-terminal do domínio quinase, $RMSD_B$ da proteína completa (considerando todos os átomos e tendo como estrutura referencial a conformação fechada) e o RMSD da hélice αC (destacada em laranja na figura 2), necessária para avaliações locais de transições, uma vez que em experimentos futuros como deleção de domínios, o RMSD da proteína completa foi alterado, invalidando-o como coordenada de reação.

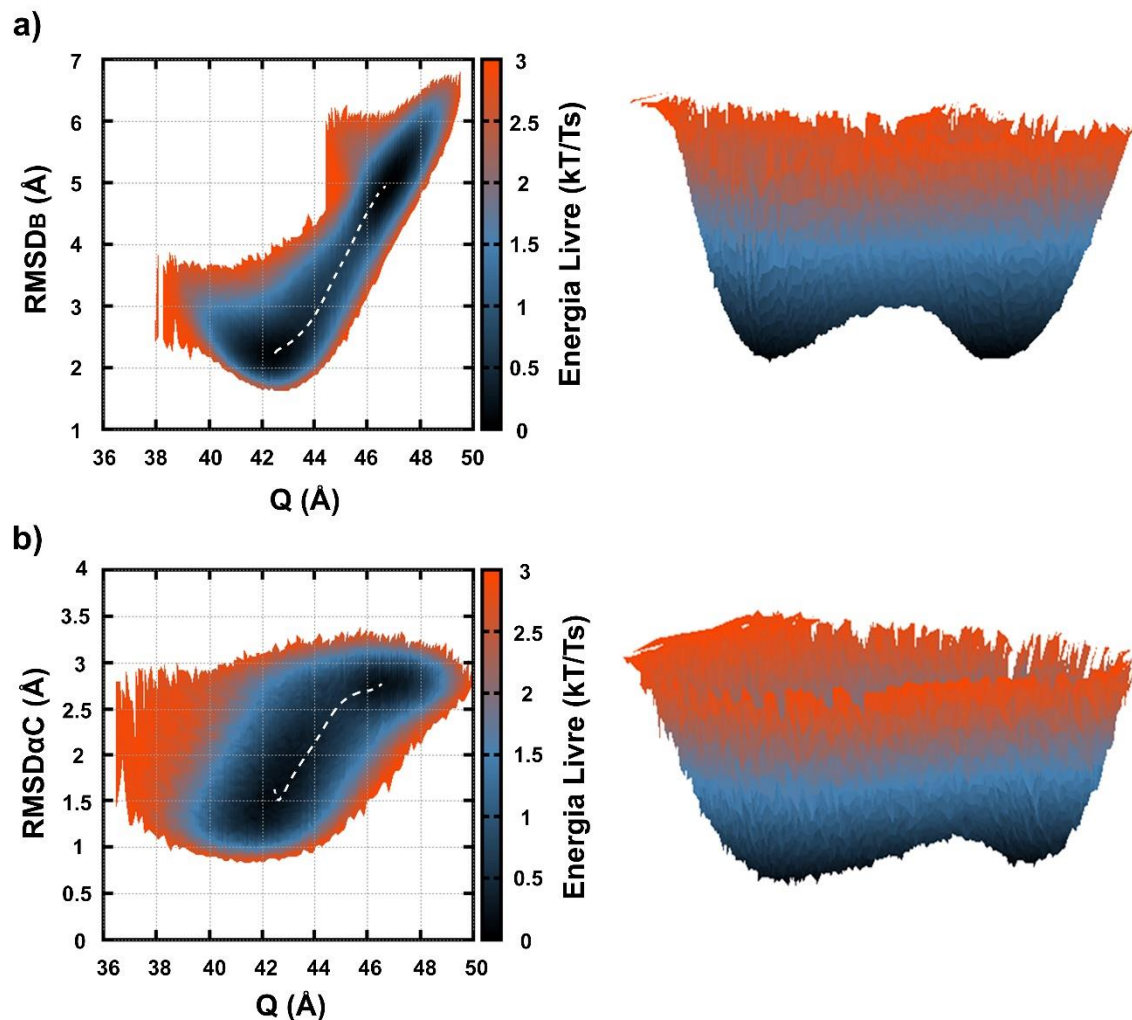
Para confirmar se essas coordenadas foram suficientes para descrever o caminho da transição entre os dois estados e garantir uma descrição mais completa do mecanismo foi utilizado, então, o Método dos Histogramas (*Weighted Histograms Analysis Method*, WHAM) (Kumar, Rosenberg, Bouzida, Swendsen, & Kollman, 1992) [Apêndice A] bidimensional e tridimensional, o qual utiliza fundamentos da física estatística, como peso estatístico e densidade de estados para calcular os perfis de Energia Livre mostrados na Figura 8.

Figura 7 - Identificação de contatos exclusivos de uma conformação. Os pontos da figura representam os contatos entre resíduos da estrutura cristalográfica fechada. No eixo das abcissas, é representado as distâncias entre esses resíduos utilizando como referência a proteína em sua forma fechada. No eixo das ordenadas, as mesmas distâncias são calculadas utilizando como referência a proteína em sua conformação aberta. Os destaques em vermelho são os resíduos em que sofrem uma variação significativa (1.5x) entre as duas estruturas e que devem ser os contatos causadores de mudanças conformacionais. Esses contatos foram selecionados como contatos de interesse e adicionados no arquivo da estrutura aberta.



Fonte: Elaborada pelo autor.

Figura 8 - Avaliação energética das mudanças conformacionais. A superfície de energia mostra uma correlação entre as coordenadas de reação RMSD e distância entre o domínio SH2 e o lóbulo C-terminal (Q) com as mudanças conformacionais, permitindo a identificação dos ensembles fechado e aberto da proteína. Para o cálculo do RMSD, foi utilizado a estrutura compacta (cadeia B - PDB 1K9A) como referência de estrutura inicial. Em (a) são utilizadas como coordenadas de reação a distância entre o domínio SH2 com Lóbulo C-terminal (Q) e o RMSD da proteína completa, correlacionando duas coordenadas globais em que utilizam a proteína completa. Em (b) é utilizado também a distância entre o domínio SH2 com Lóbulo C-terminal (Q) e o RMSD da Hélice αC , correlacionando uma coordenada global com uma local. Os perfis de energia livre são calculados pelo WHAM e as imagens são montadas utilizando o software GnuPlot (Janert, 2008).



Fonte: Elaborada pelo autor.

O modelo construído possibilitou então uma análise detalhada das transições conformacionais entre os dois estados da proteína durante as dinâmicas moleculares. Os perfis de energia livre mostram que as coordenadas de reação são suficientes para diferenciar os dois estados de mínima energia de Csk, assim como identificar o caminho (linha de mais baixa energia) entre esses.

A partir desse modelo, novas hipóteses sobre o mecanismo podem ser apontadas, permitindo sua viabilidade por experimentos *in silico* e, confrontando com resultados já obtidos *in vitro*. Nossa hipótese inicial consiste em relacionar as transições com a atividade da Csk, sugerindo que o funcionamento da regulação depende de uma ação dinâmica, em contraste com a ideia de uma conformação ativa e outra inativa.

1.5.2 Explorando a região da hélice α C.

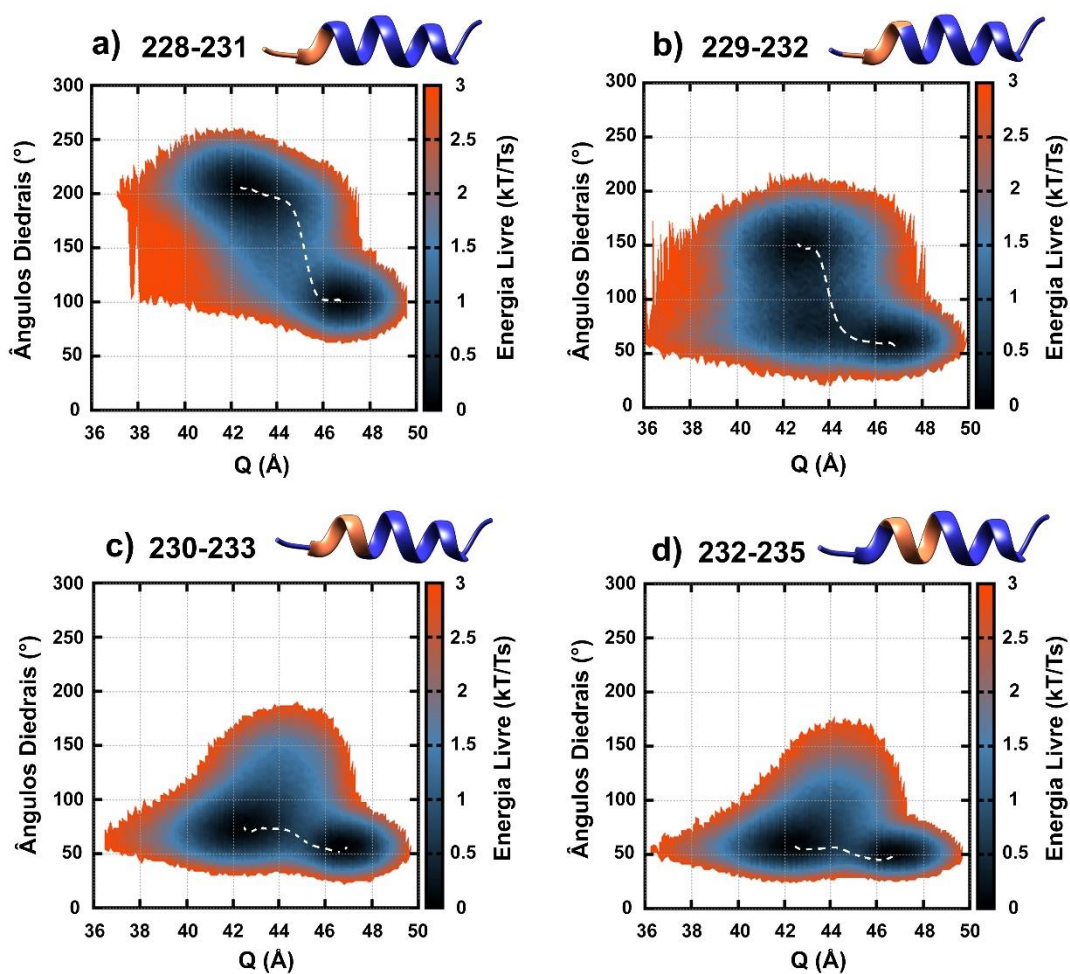
Uma região bastante promissora nos estudos envolvendo a proteína Csk é a hélice α C. Essa região sofre efeitos químicos e conformacionais quando é alterado uma outra região da proteína, no caso o domínio SH2. (Mills et al., 2007). Outro fator que comprova a importância da região são as diferenças estruturais apresentadas em seus resultados cristalográficos, indicando que a região pode apresentar algum tipo de mecanismo durante as transições, seja facilitando ou dificultando o acesso das diferentes conformações da proteína em solução.

A hélice contempla 19 resíduos (225 até 243) e para uma avaliação completa dos possíveis mecanismos da região, esses resíduos foram divididos em 16 diferentes grupos, formando os respectivos diedros. Os ângulos diedrais durante as dinâmicas moleculares foram utilizados como coordenadas de reação para verificar se há algum mecanismo envolvendo esses grupos, como por exemplo *cracking* com contatos entre alguns resíduos quebrados, relaxando a estrutura inicial e facilitando a transição, *slidding* ocorrendo apenas o rearranjo estrutural da proteína, sem quebra de interações que também facilitam as mudanças conformacionais, ou então, *back-tracking* desfazendo e refazendo as estruturas durante a mudança conformacional.

Dentre os 15 diferentes diedros analisados, quatro deles mostraram diferentes comportamentos em seus perfis de energia livre, como mostrado na Figura 9. Os dois primeiros diedros formado pelos resíduos 228 a 231 e 229 a 232, apresentados na Figura

9a e 9b respectivamente, mostram que durante as transições conformacionais, há uma correlação de movimentos, isto é, à medida em que o domínio SH2 se desloca de um estado para o outro, os diedros rotaciona juntos, indicando ser uma região promissora para receber algum tipo de mudança em que as transições conformacionais sejam afetadas.

Figura 9 - Mecanismos envolvendo a Hélice- α C. Perfil energético relacionando os movimentos globais da proteína utilizando a coordenada Q (Distância do domínio SH2 ao Lóbulo Kinase C-terminal), com movimentos locais (ângulos diedrais). (A) O perfil mostra uma correlação de movimentos entre Q e os ângulos diedrais. (B) semelhante ao comportamento de A, com variação nos valores da barreira energética, mostrando então que para essas modificações, a transição entre os estados é mais fácil devido aos valores de energias mais baixos quando comparados a (A). (C) e (D) O perfil indica os mesmos valores angulares para os dois estados conformacionais, com pequenas variações angulares nos estados de transição.



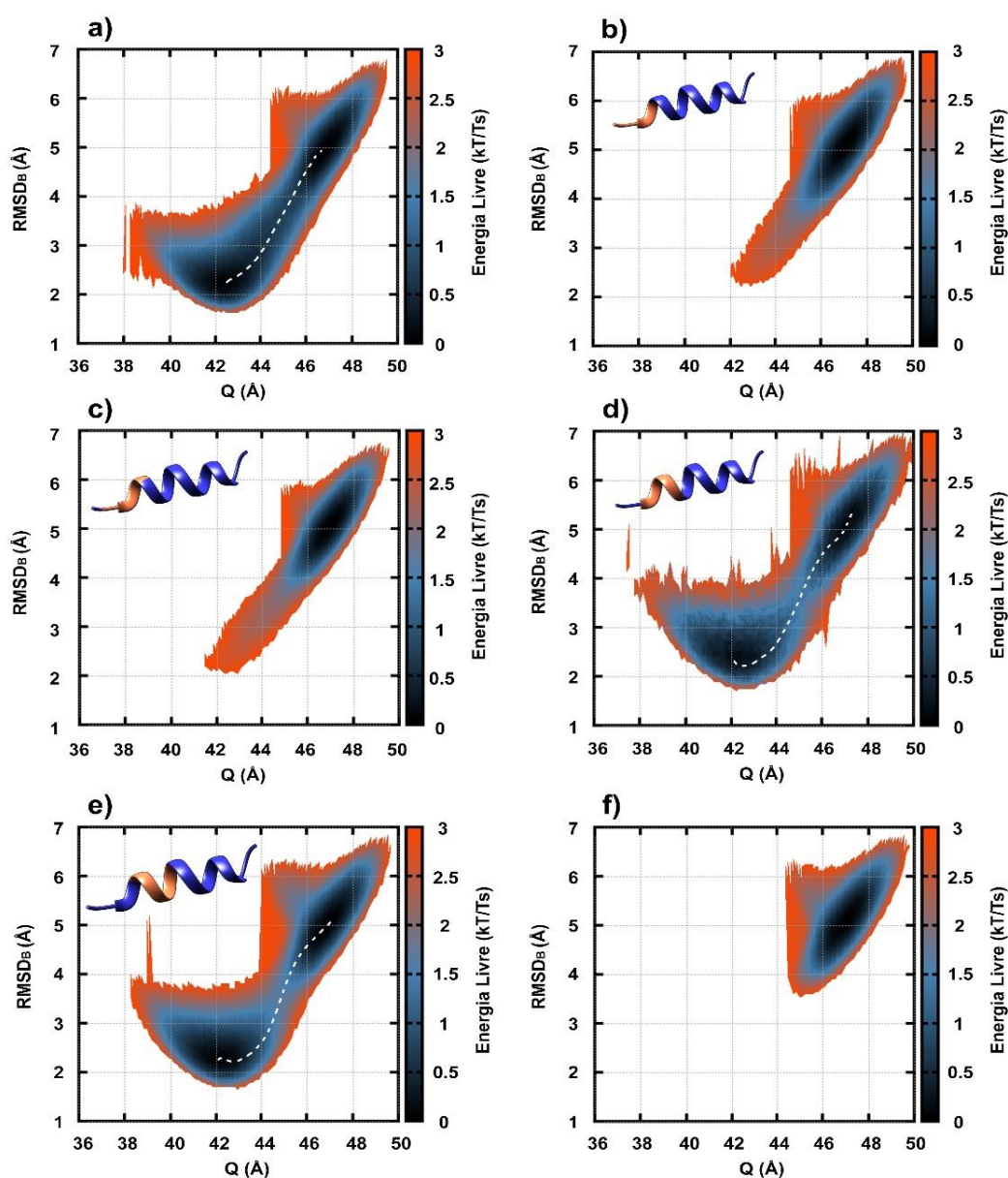
Fonte: Elaborada pelo autor

Os diedros correspondentes aos resíduos 230 a 233 e 232 a 235 mostrados na Figura 9c e 9d respectivamente mostraram que pequenas variações angulares nos diedros dessas regiões podem acontecer nos estados de transição. Esse resultado indica que alterações nesses pequenos movimentos podem ser correlacionados com o aumento de barreira energética, dificultando a passagem de um estado para outro.

Com a região alvo já fragmentada em pequenos resíduos que podem ser responsáveis pelas transições conformacionais, mudanças energéticas no pré-fator ϵ_ϕ (eq. 1) foram realizadas no arquivo de topologia da proteína Csk, de modo a enrijecer a região. Essas pequenas regiões que formam os diedros foram alteradas em seus pré-fatores energéticos ϵ_ϕ , multiplicando esses termos por uma constante $\chi=5$, dificultando seus movimentos durante as simulações computacionais.

A influência desses mecanismos para as mudanças conformacionais foram então verificadas a partir dos perfis de energia livre, fazendo uma comparação direta com os resultados obtidos para a proteína em seu estado nativo (Figura 8) e podem ser vistos na Figura 10. Vale ressaltar que quaisquer simulações com mudanças energéticas devem ser ajustadas para adequar a população de conformações. Valores de temperaturas maiores ou menores nesta modelagem não devem ser entendidas como uma mudança na termoestabilidade da proteína, assim a alteração da energia total do sistema deve ser tão somente compensada pelo novo ajuste da temperatura, sendo importante avaliar inclusive a frequência das transições.

Figura 10. Alterações energéticas em regiões alvo. Mudanças nos pré-fatores energéticos nos diedros são realizadas, inviabilizando movimentos locais. (A) Perfil energético da proteína sem modificação para comparação. (B) e (C) restrições nos movimentos dessa região fazem com que haja uma predominância do estado aberto, sendo energeticamente desfavorável para a proteína acessar o estado fechado, reduzindo consideravelmente o número de transições e indicando que os movimentos dessa região são de extrema importância para a ocorrência de mudanças conformacionais. (D) e (E) O perfil de energia altera-se de modo a aumentar a barreira energética entre os dois estados da proteína, com acesso aos dois diferentes ensembles, porém com um número menor de transições. (F) Restrições da região por completa mostra que a proteína perde os movimentos globais, acessando apenas uma conformação.



Fonte: Elaborada pelo autor.

Os resultados mostram que os movimentos da hélice αC estão correlacionados com as transições globais da proteína, sendo esses necessários para o acesso nas duas diferentes conformações. Ainda, apresentam mecanismos que podem auxiliar nas transições conformacionais, apresentando uma dinâmica no movimento apenas em estados de transição e com mesmos valores angulares nos dois estados nativos.

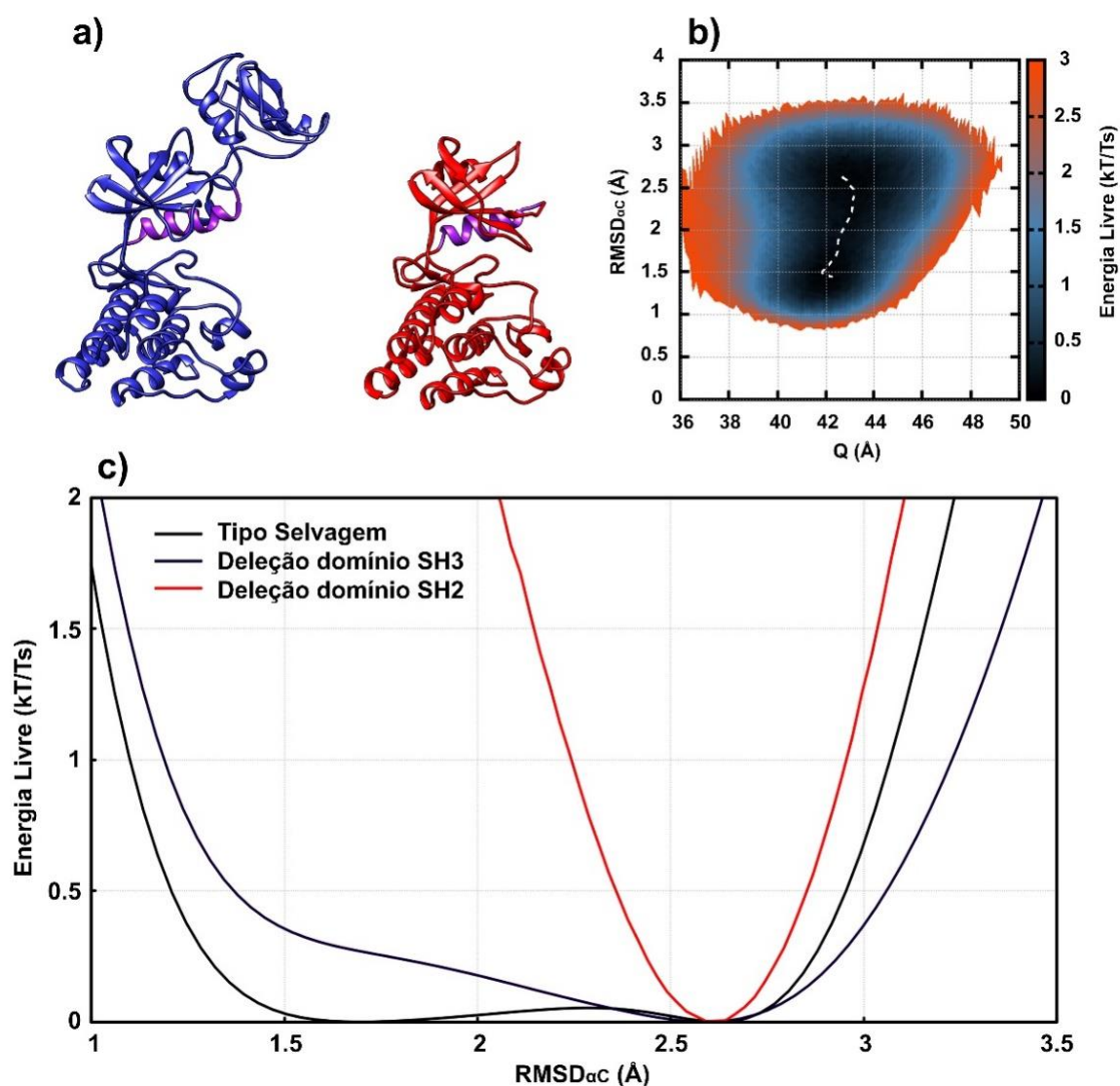
Por fim, analisando os resultados como um todo e sabendo da impossibilidade de se enrijecer regiões em proteínas reais sem mutação ou adição de outra molécula, o modelo nos permite reconhecer e mapear regiões da proteína em que fármacos poderiam atuar restringindo seus movimentos. Por consequência, o regime de flutuação da região seria modificado, dificultando ou inibindo a proteína de fazer as transições conformacionais, e possivelmente afetando sua atividade. Vale salientar ainda que, a região da hélice αC , já foi mencionada em alguns estudos como (Mills et al., 2007) e (Barkho et al., 2015), mostrando estar correlacionada com movimentos globais dos domínios proteico e suas propriedades físico-químicas (pontes dissulfetos por exemplo).

1.5.3 Deleções de domínios.

Resultados *in vitro* mostram que os movimentos do domínio SH2 estão relacionados com movimentos do domínio SH3, explicitando então que, para a transição de estados são necessários os dois domínios trabalhando em conjunto (Barkho et al., 2015). Incluso a esses resultados, a deleção do domínio SH3 da proteína Csk exibe uma redução na atividade catalítica de Csk (Barkho et al., 2015), evidenciando a necessidade da comunicação entre os domínios para sua funcionalidade.

Uma primeira verificação da sensibilidade do modelo, assim como uma primeira verificação da hipótese levantada de que a atividade está associada às transições conformacionais de Csk, utilizou o modelo de estados equiprováveis, deletando o domínio SH3 completo junto com o *linker* SH3-SH2 dos arquivos de topologia. Essa primeira análise mostrou uma conformação estável quando deletado o domínio SH3, e indicou que o modelo utilizado tem sensibilidade o suficiente para detectar a necessidade de interação cooperativa entre os domínios SH2 e SH3 para uma mudança conformacional completa, como mostrado na literatura. (Barkho et al., 2015).

Figura 11. Deleção dos domínios SH3 e SH2. (A) Ilustração das deleções. Em azul está representado a proteína Csk sem o domínio SH3 e em vermelho sem o SH3 e SH2. Em roxo é destacado a região da hélice αC . (B) Perfil energético relacionando os movimentos da hélice ($RMSD_{\alpha C}$) com o movimento do domínio Q. A coordenada Q mostra que não há mais variações de distância entre os domínios SH2 e Quinase C-terminal quando deletado o domínio SH3, ainda que a coordenada $RMSD_{\alpha C}$ não seja afetada quanto a mudanças entre os dois estados. (C) Ilustração em uma dimensão do perfil energético das deleções, sendo nesse avaliado apenas a coordenada $RMSD_{\alpha C}$, uma vez que sem o SH2 perde-se a coordenada de distâncias Q. O movimento da hélice é afetado energeticamente para a deleção de SH3. Já para deleção de SH2, não há mais transições conformacionais da hélice.



Fonte: Elaborado pelo autor.

O resultado de deleção do domínio SH3 ilustrado por seu perfil energético na Figura 10b, mostra que a coordenada de reação global Q foi afetada, restringindo a movimentação de SH2 em relação ao domínio quinase C-terminal, mostrando então a importância do domínio SH3 para a estabilidade da conformação aberta de Csk.

Apesar dos efeitos sobre os deslocamentos de SH2, a deleção do domínio SH3 mostrou que a transição conformacional não foi totalmente desfeita. A coordenada de reação local RMSD da Hélice α C tem seu perfil energético alterado, provocando apenas um aumento no custo energético, evidenciando que seus efeitos de mudanças conformacionais estão diretamente relacionados com o domínio SH2. Este resultado corrobora com experimentos *in vitro* (Mills et al., 2007), os quais mostram que uma única ponte dissulfeto no domínio SH2 de Csk afeta o lóbulo N-terminal, em específico a Hélice α C.

O segundo teste de deleção, retirando os domínios SH3 e SH2 por completo, mostrou que a coordenada RMSD da hélice α C foi alterada, de modo a não acessar o estado fechado da proteína, ficando confinado apenas no estado aberto.

Por fim, a junção dos resultados considerando ambas as coordenadas de reações e ambas as deleções, indicam que a atividade da proteína deve estar relacionada com a transição conformacional como um todo, ou seja, a proteína não deve ser totalmente funcional se tiver apenas o deslocamento do domínio SH2 sem as mudanças conformacionais da Hélice α C e vice-versa.

1.5.4 Deleções de contatos interdomínios.

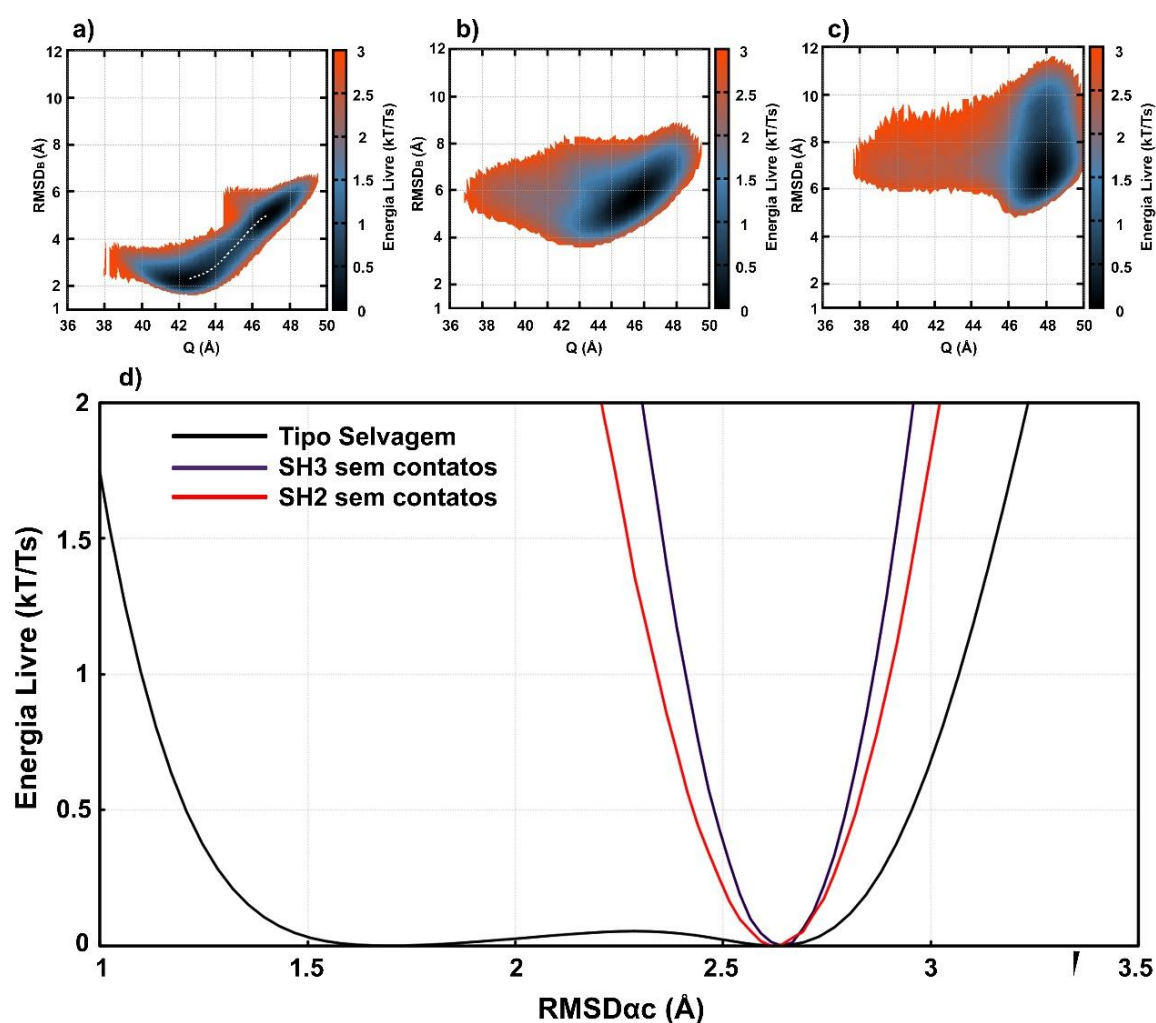
Uma vez que as deleções dos domínios interferem nos regimes de transições conformacionais, uma segunda análise contemplando os domínios de Csk é avaliar como seus contatos interferem no regime de transições, minimizando os efeitos da modelagem computacional e buscando uma possível modulação entre os domínios. Foram realizadas então dinâmicas moleculares com modelos baseados em estrutura, utilizando arquivos de topologia em que todos os contatos interdomínios de SH3 com o resto da proteína são retirados. As dinâmicas também foram realizadas para a topologia sem os contatos interdomínios SH2 com o resto da proteína.

Os resultados mostram que os efeitos de retirar o domínio SH3 e retirar seus contatos são diferentes. Quando o domínio completo é deletado, mesmo que com a energética do movimento afetada, a hélice αC ainda acessa os dois diferentes estados, como visto na sessão anterior. Entretanto, quando são retirados os seus contatos com o resto da proteína, não há mais transição da região de hélice αC como visto na Figura 12d. A restrição dos movimentos da hélice αC deve ser explicada pelo fator estérico imposto pelo domínio SH3. Quando são retirados seus contatos, SH3 restringe as mudanças conformacionais na hélice αC , ao mesmo tempo em que causa os efeitos entrópicos ao permitir um maior número de estados para o domínio SH2.

Já a deleção dos contatos interdomínios de SH2 apresentada na Figura 12b, indicaram que sua movimentação para as transições conformacionais dependem dos contatos interdomínios. Uma vez que esses contatos foram deletados, a movimentação de SH2 acontece apenas por efeitos térmicos em torno de uma das conformações. Os efeitos da deleção de contatos do domínio SH2 para com a hélice αC , mostram que seus contatos são importantes para as mudanças conformacionais locais, sendo que suas deleções causam novamente um confinamento da hélice αC no estado aberto de Csk.

Estes efeitos de restrição de movimentos e confinamento em apenas um estado, indicam que deve existir uma modulação do domínio SH3 com o SH2 e vice-versa e de ambos para com o resto da proteína, para o completo funcionamento da proteína. As correlações garantem um equilíbrio nas transições conformacionais, tanto em aspectos globais como o deslocamento do domínio SH2, quanto a aspectos locais, como as quebras de pequenas estruturas secundárias.

Figura 12 - Deleções de contatos interdomínios. (a) perfil energético das mudanças conformacionais sem alterações. (b) Perfil energético das trajetórias sem os contatos do domínio SH2 com os demais resíduos. Neste é mostrado que há mudanças conformacionais, mais a desestabilização de movimento faz com que os estados de transição entre conformação aberta e fechada seja o novo mínimo de energia. (c) Perfil energético das trajetórias sem os contatos do domínio SH3 com os demais resíduos. A coordenada de distância Q e de desvio RMSD_B mostra a perda de transição entre os dois estados conformacionais e um acesso a diversos novos estados não vistos até então, evidenciando os efeitos entrópicos e restrição de transições conformacionais. (d) perfil 2D mostrando que em ambos os casos, não há mudanças estruturais da hélice α C.



Fonte: Elaborado pelo autor.

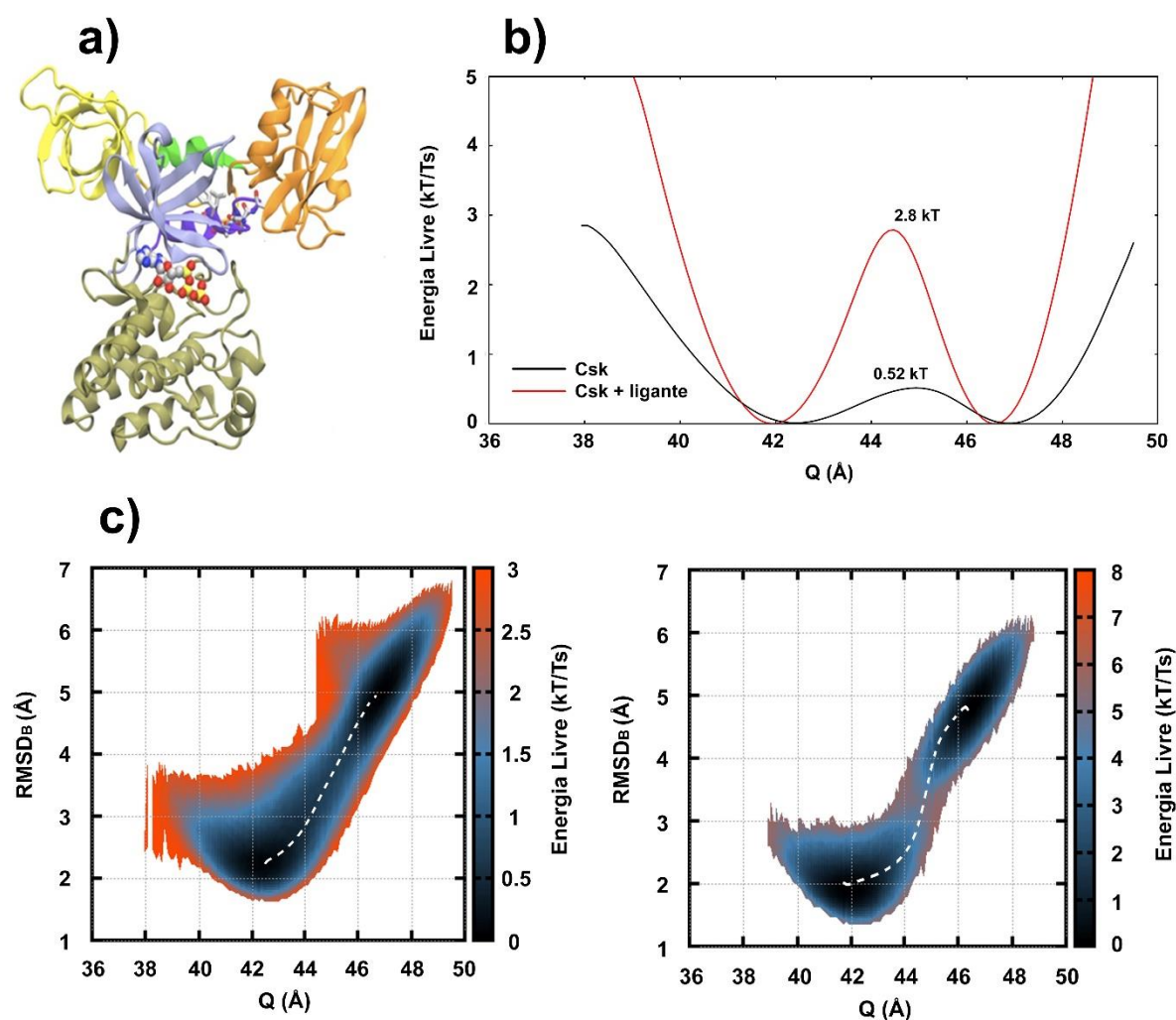
1.5.5 Modelagem de um ligante.

Como já mencionado nesse trabalho, a molécula *Staurosporine* interage com algumas das proteínas da família das SKFs (Lamers et al., 1999; Meggio et al., 1995). Aproveitando-se da homologia estrutural, a interação do ligante foi modelada, em uma primeira aproximação, utilizando-se pequenos contatos interproteicos, simulando o efeito do ligante entre os resíduos. Essa modelagem aproximada foi realizada devido ao fato de que o modelo $C\alpha$ não contabilizaria as possíveis interações eletrostáticas entre o ligante e as cadeias laterais da proteína, assim como os efeitos hidrofóbicos.

O ligante foi modelado em uma região do domínio Quinase N-terminal, como mostrado na Figura 13a. Sua presença afeta diretamente o regime de transições entre os estados. O custo energético para a transição conformacional aumenta em aproximadamente 2.3kT, afetando na dinâmica de transições, elevando o tempo para sua ocorrência, e possivelmente, na atividade da proteína, como visto na Figura 13b e 13c. Ainda que de maneira básica, esse resultado contribui para a confirmação de que as transições conformacionais da proteína ocorrem devido a uma rede de interações interdomínios e de efeitos alostéricos, e que quando correlacionados, mostram que perturbação em um domínio completo, ou até mesmo em pequenos grupos de resíduos, causam um efeito global, afetando o regime de mudanças conformacionais.

Por fim, a hipótese de que a atividade está relacionada com a transição, também foi reforçada pela investigação na presença do ligante. O fato de que a molécula modelada interfere diretamente na atividade de outras moléculas homologas a Csk, nos leva a pensar que esta pode alterar também a atividade de Csk, fazendo uma relação direta com a diminuição de transições conformacionais de Csk, como visto na Figura 13.

Figura 13 - Modelagem de um possível ligante em Csk. (A) Representação da modelagem do ligante no domínio quinase N-lóbulo. (B) Perfil energético do tipo selvagem e da proteína com o ligante. Os valores de energia indicam um aumento de aproximadamente 2.3kT na barreira de energia. (C). Na esquerda, o perfil energético da proteína sem alterações é mostrado para comparação com o resultado da direita, onde é mostrado o perfil energético da proteína com o ligante. Os valores da escala de energia são alterados devido ao aumento da barreira energética, se fazendo necessário para visualização dos mínimos de energia e do estado de transição.



Fonte: Elaborada pelo autor.

1.6 Conclusões

Foi proposto neste trabalho a avaliação das mudanças conformacionais da proteína C-terminal Src Quinase (Csk) e seus mecanismos, contribuindo assim para a elucidação de seu mecanismo de regulação, que difere de proteínas estruturalmente homólogas salvo alguns resíduos. Em uma primeira etapa, um modelo foi construído, empregando características minimalistas que utilizam de propriedades gerais do sistema para cálculos de simulações computacionais. As trajetórias das simulações permitiram o cálculo de propriedades termodinâmicas de interesse devido ao amplo espaço conformacional acessado.

Resultados experimentais da literatura foram utilizados para validação do modelo e verificação de sua sensibilidade. Uma vez que não é possível medir atividade *in silico*, junto a esses testes de validação do modelo, a hipótese que provocou a realização deste trabalho acerca de que transições conformacionais estão diretamente relacionadas à atividade da proteína foi se consolidando.

Experimentos *in vitro* mostram que o domínio SH3 da proteína tem seus movimentos correlacionados com o domínio SH2, e que sua deleção promove a diminuição de atividade da proteína (Barkho et al., 2015). Em nosso modelo, a deleção do domínio SH3 mostrou a desestabilização do domínio SH2, corroborando com os resultados de correlação dos domínios. Paralelamente à validação do modelo, a deleção acarretou em uma diminuição no regime de transições, fortificando a hipótese principal do trabalho.

Outro alvo de estudos já elucidados *in vitro*, é a deleção do domínio SH2, que resulta na diminuição quase que por completa da atividade proteica. Esse teste, também realizado em nosso modelo, mostrou que as transições conformacionais, quando verificadas por meio de uma coordenada local, são reduzidas quase que por completas, reforçando a hipótese de perda atividade com a deleção de domínio.

Outra região de interesse, com resultados já consolidados de sua importância para proteína Csk (Mills et al., 2007) é a hélice αC , região de extrema importância para a comunicação domínio SH2 com a região N-terminal do domínio Quinase. Uma análise detalhada dessa região mostra quatro grupos de resíduos responsáveis por mecanismos

como o desenovelamento local e torções que facilitam a transição conformacional. Restrições na flexibilidade desses grupos foram realizadas e mostraram que esses influenciam diretamente no regime de mudanças conformacionais, dificultando as transições. Posto que a restrição de flexibilidade envolvendo movimentos diedrais sem adição de outras macromoléculas não seja possível *in vitro*, os resultados obtidos motivam novos experimentos, como o desenho racional de fármacos que possam interagir nessa região, alterando o regime de transição de Csk e promovendo uma perda de atividade como proposto nesse trabalho.

Por fim, os resultados obtidos apontam para um modelo simplificado, porém robusto, corroborando com resultados experimentais, permitindo diversos tipos de análises e gerando novas informações a respeito da atividade da proteína assim como seu mecanismo regulador, mesmo que com hipóteses que precisam ser experimentalmente validadas.

Capitulo 2

Interação entre a proteína Grb2 e as Moléculas Cumarina e Morina

2.1 Introdução

Funções desempenhadas por uma proteína e, conseqüentemente, muitos dos processos biológicos relacionados a elas dependem de interações altamente específicas que ocorrem entre duas ou mais moléculas (Nelson & Cox, 2014). Dessa forma, muitos estudos são realizados para tentar compreender o princípio de reconhecimento e interação de uma proteína com diversas outras moléculas e macromoléculas.

Diversos são os papéis biológicos desempenhados por esse tipo de interação, podendo-se citar as interações que ocorrem entre hormônios e receptores; substratos e enzimas; antígenos e anticorpos; transportadores e solutos; ácidos nucleicos e ribossomos; entre outros (Dunn, 2010; Kessel & Ben-Tal, 2010). Portanto, o reconhecimento molecular entre duas ou mais moléculas é de extrema importância para os sistemas biológicos e é um processo controlado por interações não-covalentes específicas (Gellman, 1997). Ainda, interações entre proteína e ligante são fundamentais para quase todos os processos que ocorrem nos organismos vivos (Dunn, 2010).

As diversas funções desempenhadas por proteínas são, em grande parte, dependentes de sítios específicos de ligação para outras moléculas. Essas moléculas se ligam à proteína receptora através de uma complementariedade molecular que geralmente ocorre de forma a obter uma conformação que seja energeticamente favorável (Du, Xing, et al., 2016).

Neste trabalho, o foco foi investigar as interações da *proteína Growth Factor Receptor-Bound Protein 2* (Grb2) e as moléculas de Cumarina 1,2-Benzopirona e Morina (2',3,4',5,7-pentahidroxiflavona), ambos com interesses biológicos que serão discutidos nas próximas seções, fazendo uma ponte entre dados experimentais de interação já validados experimentalmente, e teóricos, de modo a afunilar cada processo realizado, utilizando trocas de informações como os tipos de interações e resíduos que fazem parte da interação.

2.1.1 Sinalização celular & proteína FGFR2

Organismos multicelulares podem ser considerados como sendo sociedades extremamente complexas, na qual cada indivíduo (célula) possui funções específicas e dinâmicas, visando o bem do organismo como um todo. Para que estas funções possam ser desempenhadas adequadamente, cada um dos indivíduos recebem e emitem uma grande variedade de informações, desde o contato até sinalizadores químicos sintetizados por outras células vizinhas.

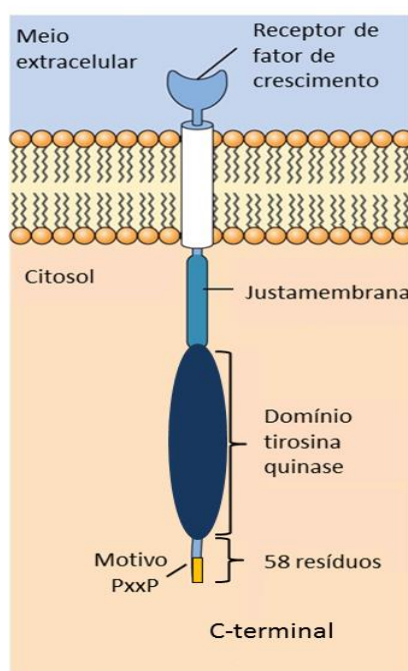
A transdução dos sinais ou sinalização celular, é, portanto, um evento-chave na resposta da célula a alterações no seu ambiente extracelular. Em eucariotos, essas sinalizações envolvem proteínas receptoras que se encontram na membrana celular e encarregam-se de receber o sinal proveniente do meio extracelular, traduzir para o meio intracelular e, assim, propagá-lo por meio de outras proteínas sinalizadoras encontradas no citosol até o núcleo. Os sinais extracelulares propriamente ditos, são interações de ligantes como hormônios, citocinas, fatores de crescimento entre outros, com a região extracelular das proteínas receptoras. (Abbas, Lichtman & Pillai, 2015).

As proteínas receptoras encontradas na membrana celular são divididas de acordo com o seu mecanismo de transdução de sinal, podendo ser proteínas receptoras acopladas a canais iônicos, as receptoras acopladas à proteína G e as proteínas receptoras associadas a enzimas (Nelson & Cox, 2014; Voet, 2013). Dentre as divisões, as proteínas receptoras que se encontram associadas a enzimas são geralmente quinases, e se apresentam como proteínas α -hélice transmembranica com um sítio de interação com o ligante na porção extracelular e o sítio catalítico na região intracelular, como é o caso da proteína receptora de fatores de crescimento tirosina quinase (*Fibroblast growth factor receptor 2* - FGFR2) representada na Figura 13. (Voet, Voet & Pratt, 2014; Rang & Dale, 2012).

FGFR2 é uma Proteína Tirosina Quinase Receptora, e sua estrutura é formada por uma região extracelular onde encontra-se o receptor de fator de crescimento, na região intracelular logo abaixo da membrana plasmática, encontra-se a região justa membrana, enquanto que subsequentemente o domínio quinase pode ser verificado. Nos últimos quinze resíduos da região C-terminal de FGFR2 encontra-se o motivo PxxP (sendo P representado pelo resíduo prolina e o duplo x por qualquer outro resíduo). Este

motivo é importante para o recrutamento de outras proteínas que fazem o reconhecimento do motivo, proporcionando sinalizações celulares (Belov, Mohammadi, & Manuscript, 2012).

Figura 14 - Representação estrutural da proteína receptora tirosina quinase. A proteína FGFR2 é composta uma região extracelular, onde encontra-se o Receptor de fator de crescimento, uma região transmembranica, uma região intracelular, que consiste de uma região justa membrana e de um domínio citoplasmático de tirosina quinase com uma sequência polipeptídica complementar de 58 aminoácidos contendo motifs PxxP responsável pela sua ligação a outras macromoléculas.



Fonte: adaptado de (Rang & Dale, 2012)

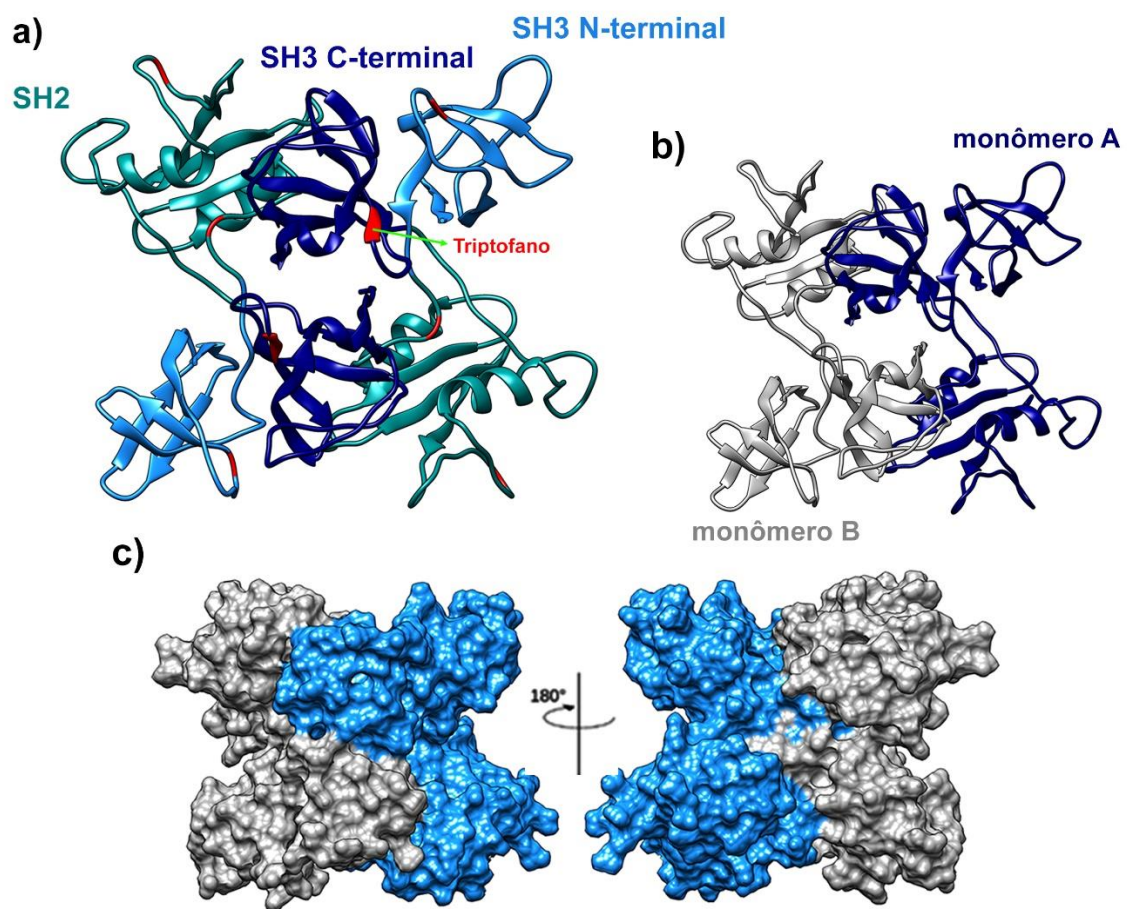
Essa proteína é responsável pela transdução de sinais extracelulares que dão início a vias de sinalização, como, por exemplo, a MAPK (*Mitogen Activated Protein Kinases*) envolvida em processos celulares fundamentais como proliferação celular, diferenciação e apoptose (Dhillon, Hagan, Rath, & Kolch, 2007). Problemas em sua funcionalidade pode acarretar em uma variedade de displasias humanas como o câncer e a má-formação fetal, bem como no desenvolvimento esquelético, devido à transdução de sinal aberrante (Ahmed et al., 2013).

Para o funcionamento e principalmente para a especificidade das vias de sinalização, proteínas de ancoragem ou adaptadoras, capazes de formar transitoriamente complexos proteicos, são fundamentais. As interações entre proteínas envolvidas na transdução de sinal com as proteínas adaptadoras, são geralmente mediadas por domínios compactos, com os N e C terminais próximos, como por exemplos os domínios SH2 (*Src homology domain 2*) e SH3 (*Src homology domain 3*), possibilitando a interação sem interferir muito em suas estruturas. Estes domínios são encontrados nas mais diversas proteínas, como por exemplo quinases, fosfatases, fatores de transcrição e proteínas do citoesqueleto (Cheng et al., 1998). Uma importante proteína adaptadora que interage com a via mediada pela proteína receptora FGFR2 é a Grb2 (*Growth Factor Receptor-bound 2*), a qual foi alvo de nossos estudos.

2.1.2 Growth Factor Receptor-bound 2 - GRB2.

A Grb2 é uma proteína adaptadora citosólica, ubiquamente expressa em eucariotos. Contendo 217 aminoácidos e massa molecular de 25 kDa, é constituída por um domínio SH2 ligado em dois domínios SH3, (Maignan et al., 1995), como pode ser visualizado na Figura 15. É uma proteína essencial para a transdução de sinais intracelulares através de interações com outras proteínas em regiões contendo fosfotirosina quando (interação com o domínio SH2) ou sequências ricas em prolinas (interação com os domínios SH3). Essas especificidade de interações fazem de Grb2 uma proteína adaptadora envolvida em processos complexos como a morfogênese epitelial, angiogênese e vasculogênese, além de estar relacionada com a prevenção de tumores e metástases (Giubellino, Burke, & Bottaro, 2008).

Figura 15 - Proteína Grb2 em sua forma dimérica. Estrutura tridimensional da proteína Grb2 em sua forma dimérica retirada PDB (PDBid 1GRI). (A) A proteína Grb2 está colorida diferenciando seus três domínios em cada monômero. São destacados também os 5 resíduos de triptofanos em vermelho, resíduos estes utilizados nas próximas seções como parâmetro para as avaliações computacionais. (B) Grb2 colorida por seus dois monômeros, mostrando que a interface dos monômeros para a formação do dímero se faz através de resíduos dos domínios SH3 C-terminal e SH2. (C) Superfície molecular colorida para diferenciar os monômeros de Grb2. A utilização de sua forma dimérica se explica pelo fato de que os experimentos *in vitro* de interação foram realizados utilizando supostamente dímeros.

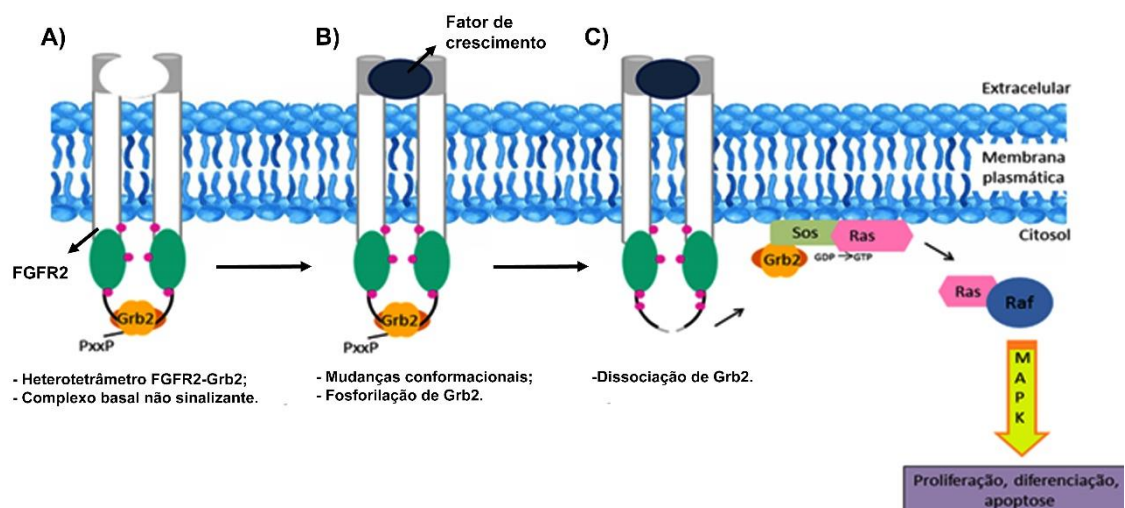


Fonte: Elaborado pelo autor.
PDBid: 1GRI

A literatura sugere que a Grb2 está em um constante equilíbrio entre suas formas dimérica e monomérica, sendo cada uma delas associada a diferentes funções (Ahmed et al., 2015). Estudos recentes mostraram que a interação da proteína adaptadora Grb2 com a proteína quinase FGFR2 ocorre na ausência de estímulos extracelulares (C. C. Lin

et al., 2012). Este mecanismo está representado na Figura 16, onde FGFR2 sofre fosforilação antes de estímulo extracelular que resulta em um chamado “nível de fosforilação basal” que não é capaz de ativar uma via de sinalização, mas sim, um estado fosforilado e não sinalizante. Ainda, neste estado Grb2 é capaz de ligar-se via domínio C-SH3 na região C terminal de FGFR2 formando um complexo basal FGFR2-Grb2. Quando fatores de crescimento se ligam na porção extracelular de FGFR2, mudanças conformacionais e fosforilações são induzidas, levando Grb2 a se dissociar de FGFR2 (Ahmed et al., 2013).

Figura 16 - Representação do mecanismo de regulação da proteína quinase FGFR2 através da proteína adaptadora Grb2. A proteína FGFR2 é representada mimeticamente por barras representando sua região citoplasmática e com seu domínio quinase representado pela elipse colorida em ciano. (A) A proteína Grb2 dimérica interage com FGFR2 através dos seus domínios SH3 C-terminais, formando um heterotetrâmero na proporção 2:2; (B) Fatores de crescimento representados pelas elipses em azul escuro ligam-se na porção extracelular de FGFR2, induzindo mudanças conformacionais em FGFR2 e a fosforilação de Grb2; (C) uma vez fosforilada, Grb2 se dissocia do domínio quinase por repulsões eletrostáticas e FGFR2 encontra-se livre para recrutar proteínas do citosol para iniciar cascatas de sinalizações.



Fonte: Adaptado de (C. C. Lin et al., 2012) por Karoline Sanches.

Após a dissociação, outras proteínas do citosol podem ser recrutadas por FGFR2 dando início a complexos de sinalização primários, como os da via MAPK, até que ocorra

a desfosforilação de Grb2 mediada pela fosfatase SHP2 e o complexo Grb2-FGFR2 seja restabelecido, voltando ao seu estado ocioso. (C. C. Lin et al., 2012).

Este ciclo envolvendo fosforilação e desfosforilação entre Grb2 e FGFR2 proporciona a regulação da atividade quinase da proteína FGFR2. Uma vez que fatores de crescimento se ligam na receptora FGFR2 esta torna-se ativa para recrutar proteínas e iniciar cascatas de sinalização que, na ausência de regulação, podem causar danos no organismo como o crescimento celular descontrolado, podendo ocasionar o desenvolvimento de tumores (Ahmed et al., 2013).

Nesse contexto, Grb2 é promovida de uma simples adaptadora para um importante regulador de processos envolvendo FGFR2 (Ahmed et al., 2013), tornando-se então um alvo potencial para o estudo de interações com moléculas de interesse farmacêutico que possam intervir em processos de sinalização celular descontrolada, que levam ao desenvolvimento tumoral.

2.1.3 A molécula de Cumarina

Cumarinas (do francês *Coumarou*) são membros da família dos compostos benzopirenos, que consistem em um anel benzeno junto a uma pirona (Mirunalini e Krishnaveni, 2011). Atualmente, já foram isoladas mais de 3.400 cumarinas sendo que seu sistema de anéis constitui uma estrutura útil na síntese de compostos heterocíclicos e fármacos (Audisio et al., 2010). A classificação das Cumarinas ocorre de acordo com as substituições nos anéis benzeno e pirano (Venugopala, Rashmi, & Odhav, 2013), e podem ser classificadas em Cumarinas simples, furanocumarinas, piranocumarinas, e biscumarinas como mostrado na Figura 16.

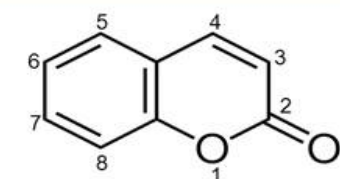
São classificadas como metabólitos secundários, ou seja, desempenham importante papel na defesa dos vegetais (Montagner, 2007). Elas foram primeiramente encontradas em sementes de *Dipteryx odorata* selvagem, uma árvore popularmente conhecida como Cumaru, nativa no Brasil, Colômbia, Peru e Guiana (Venugopala et al., 2013). Essas moléculas são amplamente distribuídas nos frutos, sementes, raízes e folhas de plantas, principalmente em angiospermas e dicotiledôneas como em cenouras, aipo, salsa, chicória e frutas como toranja (*grapefruit*). Podem ser

encontradas também em óleos de cássia (canela-chinesa), óleo de casca de canela (Caracelli et al., 2015; Montagner, 2007).

Durante os últimos anos, a procura por medicamentos de origem vegetal tem gerado um grande interesse dos pesquisadores nas Cumarinas e seus derivados devido as suas potentes atividades farmacológicas. Suas propriedades como moléculas biológicas vão desde antibactericida, anti-inflamatório, anticoagulante, antitrombótico, vasodilatador até atividade anticancerígena comprovada (Marshall et al., 1994; Thakur, Singla, & Jaitak, 2015).

Figura 17 - Classificação das moléculas de Cumarinas. As moléculas de Cumarina são classificadas de acordo com as substituições dos anéis benzenos. Aqui são exemplificados os três diferentes tipos de Cumarinas com suas respectivas estruturas básicas.

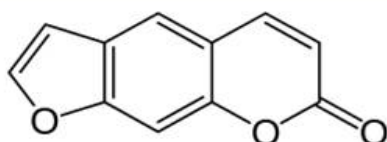
Cumarina simples



1,2 – benzopirona

- Apresentam substituições nos anéis benzeno e pirano;

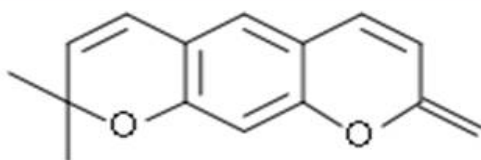
Furanocumarinas



6,7 – furanocumarina

- Apresentam anel furano junto a estrutura de cumarina simples;

Piranocumarinas



6,7 – piranocumarina

- Apresentam anel pirano junto a estrutura de cumarina simples;

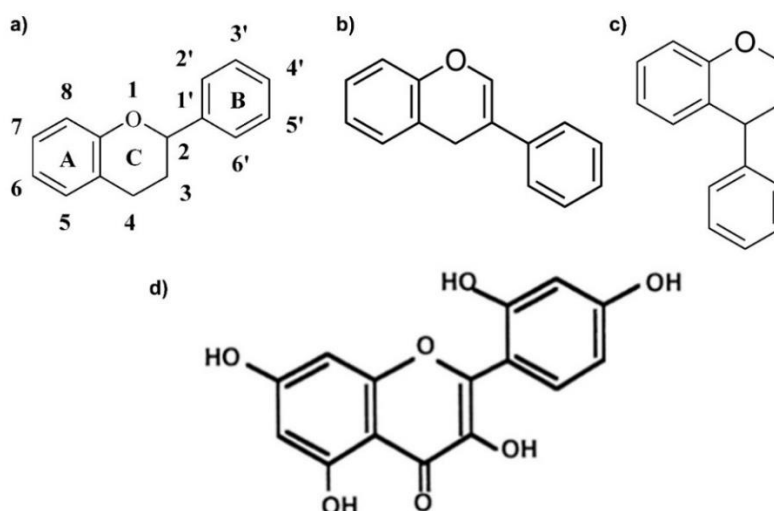
Fonte: adaptado de (Caracelli et al., 2015; Lacy, 2004)

Os estudos utilizando Cumarinas têm aumentado uma vez que o seu desempenho em atividades biológicas é bastante amplo. Foi verificado por *Mohler e colaboradores* (1994) a atividade anticarcinogênica da Cumarina em células prostáticas, que provocam a repressão de câncer em ratos (Marshall et al., 1994). López-González (2004) verificou a eficácia da Cumarina *in vitro* na indução de apoptose de adenocarcinoma pulmonar (Lopez-Gonzalez et al., 2004). A Cumarina também apresentou atividade inibitória semelhante ao efeito do tratamento quimioterápico tamoxifeno em câncer de mama (Musa, Cooperwood, & Khan, 2008). Estas, entre outras, revelam o grande potencial da molécula Cumarina, o que torna seu estudo fundamental em áreas multidisciplinares.

2.1.4 A molécula de Morina

A Morina (2',3,4',5,7-pentahidroxifavona) é um flavonol, subgrupo dos flavonoides, isolado a partir de várias espécies vegetais, como, por exemplo, agrião, laranja de osage (mesma família das amoras), folhas de goiabeira e folhas de oliveira (Pereira et al, 2015). Sua estrutura apresenta hidroxilas nas posições 5 e 7 do anel benzênico A e nas posições 2' e 4' do anel benzênico B, além de uma hidroxila na posição 3 do anel pirona, de acordo com a composição básica de um flavonol da Figura 18a.

Figura 18 – Classes e estruturas gerais dos flavonoides e flavonoide Morina. (a) estrutura geral dos flavonoides onde o anel B é ligado ao carbono 2 do anel Pirona C. (b) isoflavonoides com o anel B ligado ao carbono 3 do anel pirona. (c) neoflavonoides com o anel B ligado ao carbono 4 do anel pirona. (d) O flavonoide Morina (2',3,4',5,7- pentahidroxi flavona) apresenta hidroxilas nas posições 5 e 7 do anel benzênico A, nas posições 6' e 4' do anel benzênico B e na posição 3 do anel pirona. A numeração de cada átomo, e as letras que representam os anéis em (A) são igualmente representadas para (b) e (c).



Fonte: adaptado de (Kawabata et al, 1999).

Estudos apontam que o flavonoide Morina, inibe a proliferação celular em diversos tipos de células carcinogênicas, inibindo o ciclo celular em tumores como carcinoma oral e em hepatocarcinoma humanos. A Morina induz morte celular por apoptose em linhagens de células de câncer de pulmão do tipo não pequenas (H460), dependendo da concentração e tempo de tratamento, alterando a função mitocondrial (Pereira et al, 2015). Esse tipo de câncer é associado à atividade aberrante na via de sinalização MAPK, que tem sua regulação mediada pela proteína quinase receptora FGFR2 e que por sua vez é regulada pela proteína adaptadora Grb2 estudada (Timsah et al. 2015). Estima-se que 80 a 85% dos cânceres de pulmão são do tipo dessas células não pequenas. É mostrado ainda que, Morina possui baixa toxicidade para o organismo, tornando-a um agente em potencial de desenvolvimento de fármacos (Yazici et al, 2012) (Said et al, 2014), porém mais testes devem ser realizados com o objetivo de elucidar as vias de indução por apoptose promovida pela Morina nessa linhagem neoplásica (Pereira et al, 2015).

2.2 Motivações e Objetivos

A proteína Grb2 atua como reguladora da atividade quinase da proteína FGFR2, além de sua função de adaptadora, que é indispensável em processos de sinalização celular. Dessa maneira, Grb2 é um alvo em potencial para o estudo de interações com pequenas moléculas de interesse biológico que podem intervir em processos de sinalização celular descontrolada, que levam ao desenvolvimento tumoral.

Estudos revelam a ocorrência de interação entre a proteína Grb2 e as moléculas Cumarina e Morina, na qual a interação predominante bem como uma pré modelagem de como o ligante se liga ao sítio foram vistos experimentalmente. Sendo assim, torna-se essencial uma abordagem computacional para verificação do sítio de interação afim de complementar a compreensão de seu mecanismo de interação.

As moléculas Cumarina e Morina foram escolhidas por suas características como: propriedades farmacológicas como antibactericida, anti-inflamatório, anticoagulante, antitrombótica, vasodilatador e atividade anticancerígena comprovada, além do baixo custo e fácil acesso. Uma vez que encontradas os sítios de interação para essas estruturas de composições estruturais básicas, modificações em trabalhos futuros podem ser realizados, de modo a aumentar a afinidade das moléculas dependendo do objetivo a ser estudado.

Com base nas informações já validadas experimentalmente, o objetivo do trabalho foi caracterizar os possíveis sítios de interação da proteína Grb2 com as moléculas de Cumarina e Morina, fazendo a junção teórico-experimental, por meio de das técnicas computacionais de *docking* e dinâmica molecular, baseando-se em informações experimentais já estabelecidas, tais como, espectros de fluorescência, epítomos obtidos por Ressonância Magnética Nuclear (RMN) e perfis termodinâmicos da interação.

Uma vez que já é conhecida a ocorrência da interação entre os ligantes, não foi realizado um tratamento inicial baseado em *high throughput*, testando a interação outras moléculas por meio de banco de dados.

2.3 Fundamentação Teórica

2.3.1 Resultados Experimentais

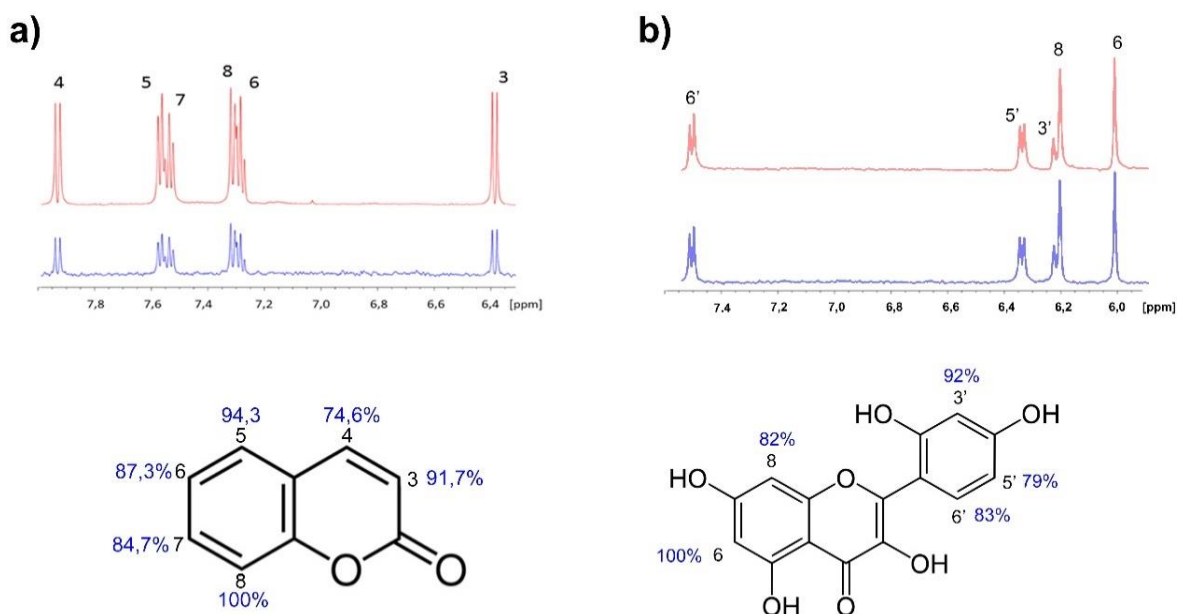
Experimentos de RMN-STD (Ressonância Magnética Nuclear - *Saturation Transfer Difference*) e espectroscopia de fluorescência foram previamente realizados por colaboradores para a verificação de interação entre a proteína Grb2 e os ligantes Cumarina e Morina. O STD consiste da diferença de dois experimentos, o *off-resonance* e o *on-resonance*. No primeiro experimento (*on-resonance*), os hidrogênios da proteína são seletivamente saturados via pulsos de radiofrequência. Os pulsos de radiofrequência são aplicados numa frequência a qual abrange somente os sinais da proteína. Os hidrogênios saturados da proteína propagam a saturação para outros hidrogênios via rede de interações dipolares intermolecular H-H (Angulo & Nieto, 2011; Martins & Jocelyne, 2007; Meyer & Peters, 2003), por difusão de spin. Este processo está relacionado com a massa molar da proteína, por isso quanto maior for o tamanho da proteína, mais eficiente será a difusão de spin. Os ligantes que interagem com a proteína recebem magnetização, se dissociam e voltam para a solução, onde seu estado saturado persiste devido ao pequeno tempo de relaxação longitudinal do mesmo (Calabrò et al., 2005; Fielding, 2007).

O segundo experimento *off-resonance*, é o experimento de referência, obtido a partir de uma sequência de pulsos fora da região espectral, onde não há sinal da proteína nem do ligante. Os experimentos *off-resonance* e *on-resonance* são subtraídos e, com a diferença dos dois experimentos têm-se apenas os sinais do ligante que recebeu magnetização da proteína, resultado da interação proteína-ligante. Quanto maior esse sinal, mais específica é a interação (Angulo & Nieto, 2011).

Os resultados de RMN-STD apresentados na Figura 19 mostram que está ocorrendo interação entre a proteína Grb2 e os ligantes Cumarina e Morina. Durante o experimento, o ligante interagiu com a proteína, recebendo saturação da mesma. Após a dissociação do complexo, devido a característica moderada da interação, a transferência de saturação pode ser observada através do espectro obtido. A região do ligante que está mais próxima à proteína durante a interação é aquela cuja intensidade

do espectro de diferença é maior, concedendo assim informações de como o ligante deve estar encaixado na macromolécula.

Figura 19 - Experimento de STD. Espectros de *off resonance* (vermelho) e *difference* (azul) evidenciando a transferência de saturação e consequentemente a interação existente entre as moléculas Cumarina e Morina com a proteína Grb2. Os números são referentes ao assinalamento dos hidrogênios da Cumarina e Morina respectivamente. As estruturas dos ligantes encontram-se abaixo aos seus respectivos espectros e marcados com os assinalamentos juntamente com as porcentagens referentes à proximidade de contato dos hidrogênios com a proteína Grb2. Em (A) é representado o espectro e estrutura da molécula de Cumarina e em (B) é representado o espectro e estrutura da molécula de Morina.



Fonte: Experimentos realizados pelo grupo de Biofísica Molecular do professor Fernando Alves de Melo, pelos alunos Karoline Sanches e Paulo Henrique da Silva.

Complementar ao experimento de RMN, que evidenciou as interações Proteína-Ligante a espectroscopia de fluorescência foi utilizada para caracterizar os parâmetros termodinâmicos da ligação. Devido ao fato de que a proteína Grb2 apresenta 5 triptofanos em seus monômeros, divididos entre seus três domínios, esta técnica foi realizada acompanhando a supressão da intensidade de fluorescência desses resíduos

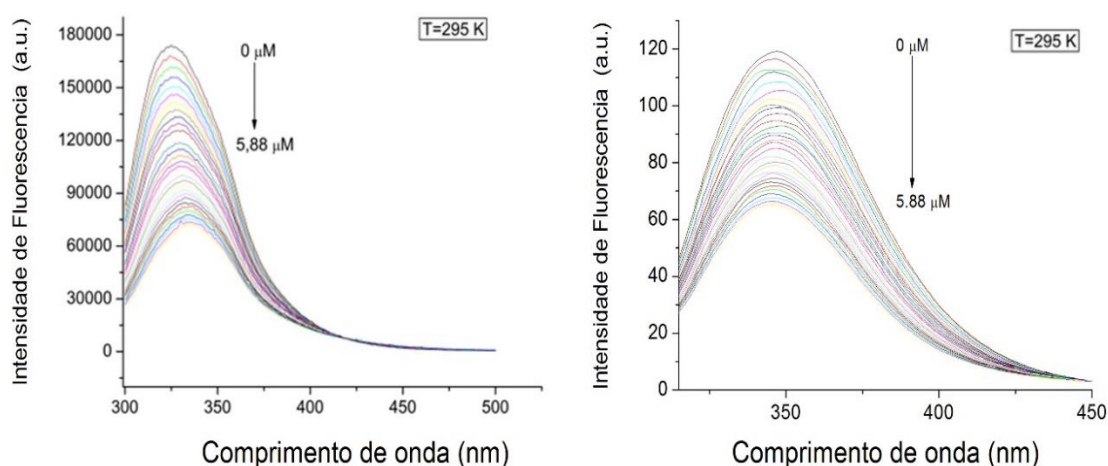
(Menezes, Fedorov, Baleizão, Valeur, & Berberan-Santos, 2013). Diversos tipos de interações moleculares podem ocasionar a supressão, dentre elas estão as reações de estado excitado, rearranjos moleculares, transferência de energia, formação de complexos no estado fundamental e supressão colisional ou dinâmica (Lakowicz, 2011). Importantes informações sobre sistemas bioquímicos têm sido evidenciadas através da supressão de fluorescência, principalmente sobre interações moleculares, que resultam na supressão da intensidade de fluorescência dos fluoróforos presentes nas macromoléculas.

Os resultados mostraram, através do espectro de emissão de fluorescência da proteína Grb2 apresentado na Figura 20, que ocorreu a supressão na intensidade do sinal de fluorescência com o aumento da concentração de Cumarina e Morina. Os resíduos de triptofano da proteína foram excitados em 290 nm e a emissão de fluorescência foi verificada entre 300 e 500 nm. A supressão na intensidade de emissão de fluorescência indica que um ou mais resíduos de triptofano pertencente a proteína está sendo afetado pela interação entre as moléculas, reduzindo assim o número de sítios totais à apenas aqueles em que contém resíduos de triptofanos.

A emissão de fluorescência no comprimento de onda em 325nm com um pequeno deslocamento durante a supressão para 330nm para a interação Grb2 e Cumarina, indicou que o resíduo que sofre a supressão, no caso o triptofano, deve estar localizado dentro de alguma cavidade e não exposto ao solvente, com seus graus de liberdade intermediários.

O experimento de fluorescência foi realizado em três diferentes temperaturas (290K, 295K e 300K) todavia esses dados não foram mostrados porquê os experimentos computacionais foram realizados na temperatura intermediária (295K). A utilização das três temperaturas no experimento *in vitro* possibilitam um tratamento mais sofisticado do espectro obtendo assim parâmetros como a constante de ligação $K_B = 1,9 \cdot 10^5 \text{ M}^{-1}$ e número de ligantes por proteína $n=1$ para a Cumarina e $K_B = 1,5 \cdot 10^5 \text{ M}^{-1}$ com $n=1$ para a Morina.

Figura 20 - Espectro de emissão de fluorescência da proteína Grb2. O primeiro espectro (cor preta) é obtido na ausência de ligantes (0 μM). Na presença dos ligantes em diferentes concentrações, é visualizado a supressão da intensidade de fluorescência, indicando que está ocorrendo a interação entre a macromolécula e o ligante. Devido a excitação ser realizado em $\lambda_{\text{exc}}=290\text{ nm}$, o fluoróforo suprimido é um ou mais triptofanos pertencentes a macromolécula. Na Esquerda, o espectro é referente à molécula de Cumarina na temperatura de 295 K. À direita, o espectro é referente à molécula de Morina na temperatura de 295 K. A diferença de medidas de intensidades é devido ao aparelho utilizado para medida. Vale ressaltar ainda, que os experimentos foram realizados em três diferentes temperaturas (290K, 295K e 300K) no entanto só estão sendo mostrados os espectros obtidos a 295K.



Fonte: Experimentos realizados pelo grupo de Biofísica Molecular do professor Fernando Alves de Melo, pelos alunos Karoline Sanches e Paulo Henrique da Silva.

Os valores obtidos para a constante de associação indicam que a interação entre a proteína Grb2 e os ligantes é moderada e com ordem de grandeza 10^5 M^{-1} (Ding, Zhao, Huang, Sun, & Zhang, 2009) onde um ligante interage com uma proteína, sendo a proporção proteína-ligante de 1:1.

Uma vez que o interesse nas moléculas Cumarina e Morina são suas propriedades farmacológicas é conveniente que a interação seja moderada. Este tipo de interação permite que a proteína Grb2 complexada com os ligantes encontrem moléculas com maior especificidade dentro da célula, ocorrendo a dissociação, permitindo assim que os ligantes possam ser entregues em vias específicas no citosol.

Por fim, os parâmetros termodinâmicos de interesse como Entalpia, Entropia e Energia Livre de Gibbs foram calculados através dos resultados obtidos por fluorescência e são mostrados na Figura 21. A Energia livre de Gibbs pode ser associada à constante de ligação por meio da Eq.3

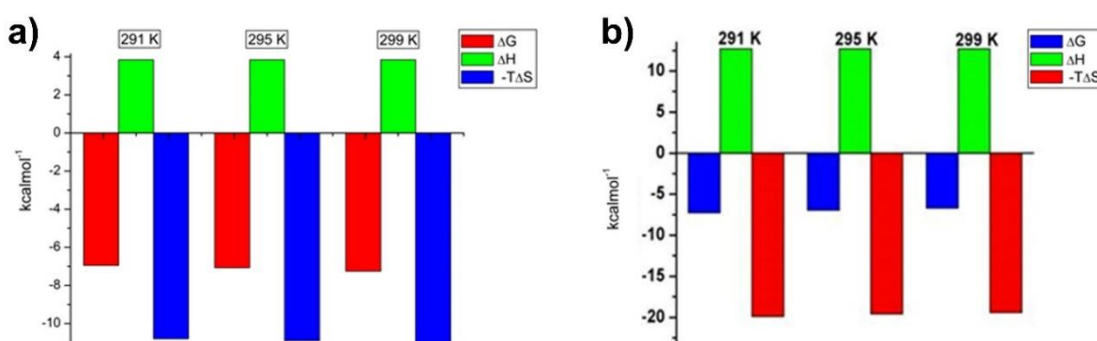
$$\Delta G = -RT\ln(K_b) \quad (\text{Eq. 3})$$

Sendo R é a constante universal dos gases, T é a temperatura e K_b é a constante de ligação. Para sistemas os quais a variação de entalpia não é significativa, a equação de van't Hoff (Eq. 4) pode ser aplicada (Ladbury & Chowdhry, 1996)

$$-R\ln(K_b) = \frac{\Delta H}{T} - \Delta S \quad (\text{Eq. 4})$$

As contribuições entálpicas estão associadas ao calor do sistema, a incluir eventos endotérmicos e exotérmicos, formação de interações não covalentes entre biomoléculas bem como o calor associado a outras formas de equilíbrio (Ladbury & Chowdhry, 1996). Já as contribuições entrópicas estão associadas a mudanças conformacionais, por exemplo, alterações na distribuição de moléculas de água em cavidades superficiais a partir de interações biomoleculares. A contribuição entrópica é dominante em muitas interações proteína-ligante, nas quais as moléculas de água são liberadas de cavidades hidrofóbicas (Ladbury & Chowdhry, 1996). Em proteínas, interações hidrofóbicas ocorrem tipicamente entre aminoácidos como leucina, isoleucina, alanina, valina e fenilalanina (Voet, Voet & Pratt, 2014).

Figura 21 – Perfil termodinâmico obtido por fluorescência. A) Perfil termodinâmico para a interação entre Grb2 e a molécula de Cumarina, as contribuições entrópicas são coloridas em azul, entálpicas em verde e a energia livre de Gibbs em vermelho. B) Perfil termodinâmico para a interação entre Grb2 e a molécula de Morina, as contribuições entrópicas são coloridas em vermelho, entálpicas em verde e a energia livre de Gibbs em azul. Os resultados aqui estão mostrados utilizando as três temperaturas em que o experimento de fluorescência foi realizado.



Fonte: Experimentos realizados pelo grupo de Biofísica Molecular do professor Fernando Alves de Melo, pelos alunos Karoline Sanches e Paulo Henrique da Silva.

Os resultados mostram que a energia livre de Gibbs obtida é negativa ($\Delta G < 0$) indicando uma reação espontânea para as interações. A maior contribuição entrópica revela que esta é mais favorável ao sistema que as contribuições entálpicas. Este é o padrão esperado para o caso de interações com efeitos do tipo hidrofóbicos, acarretando em mudanças conformacionais, expondo bolsões hidrofóbicos e liberando moléculas de água da camada de solvatação para o meio, a fim de que as moléculas sejam capazes de se ligar à proteína, escondendo-se do meio aquoso (Ladbury & Chowdhry, 1996). Este processo faz com que a entropia do sistema aumente, ou seja, esta contribuição termodinâmica é favorável para que as interações Grb2-Cumarina e Grb2-Morina ocorra.

Baseado em todas essas informações experimentais, os parâmetros para realizar o *docking* molecular foram a presença do triptofano no bolsão de interação, a predominância de efeitos hidrofóbicos na interação proteína-ligante e os epítomos de interação que mostraram como os ligantes devem estar posicionados nos possíveis bolsões.

2.3.2 Docking molecular.

Diversos substratos e inibidores, candidatos a novos fármacos, são tipicamente descobertos pela triagem *in vitro*. Apesar disso, este processo pode ser considerado muito dispendioso quanto ao tempo e custo associados. Se uma estrutura tridimensional do alvo em estudo encontra-se disponível, a simulação através de *docking* molecular pode ser uma ferramenta muito útil no estudo e descoberta de novos compostos que possam atuar como fármacos (Pierri, Parisi, & Porcelli, 2010). Assim, o *docking* molecular é uma técnica computacional que visa identificar encaixes tridimensional entre moléculas, como por exemplo interações proteína-proteína, proteína-DNA ou proteína-ligante como é o caso deste trabalho.

Originalmente, o *docking* surge como uma ferramenta que calcula a forma preferencial com que pequenos ligantes se acomodam em sítios específicos de macromoléculas biológicas, estimando suas afinidades de ligação (Yuriev, Agostino, & Ramsland, 2011). Considerada como uma das metodologias pioneiras nos anos 80 (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982), o sistema de *docking* molecular proteína-ligante consiste de duas etapas: um algoritmo de busca conformacional envolvendo os graus de liberdade do ligante (translacional, rotacional) e uma função *score* simplificada, atribuindo critérios físicos como energias de interações que classifique as prováveis posições espaciais para um composto ou mais provenientes de uma biblioteca (Rognan, 2011).

Para resolver problemas envolvendo funções *score* simplificadas, alguns pacotes de *docking* trazem funções *rescore* que utilizam abordagens mais completas para o cálculo de interação e energias, como o MM/GBSA (*Molecular Dynamics, Generalized Born and Solvent Accessibility*) (Graves et al., 2008), que utiliza trajetórias de dinâmica molecular (apêndice B) com solvente implícito para calcular as estimativas da energia de interação entre pequenos ligantes e alvos macromoleculares, como será apresentado na seção destinada aos métodos computacionais para *docking* molecular.

2.4 Modelos e métodos

2.4.1 Preparação da macromolécula

Informações experimentais, conforme mencionado anteriormente, mostram que há uma interação entre a proteína Grb2 e os ligantes Cumarina e Morina, assim é necessário que essas moléculas sejam preparadas computacionalmente para a realização do *docking*. Inicialmente, a estrutura tridimensional das moléculas de Cumarina e Morina foram retiradas do banco de dados PubChem (Bolton, Wang, Thiessen, & Bryant, 2008) com os códigos CID 323 e 5281670 respectivamente.

Para o receptor, grande parte das estruturas foram resolvidas para macromolécula em sua forma monomérica e em domínios específicos³, todavia os resultados experimentais obtidos em colaboração são realizados com a proteína em sua forma dimérica, sem a evidência de interação em domínios específicos. Assim, foi realizada uma busca por uma estrutura cristalográfica na forma dimérica, evitando a inviabilidade de uso imediato devido à obstrução de cavidades, indisponibilidade da interface ou outras razões adversas. A estrutura utilizada durante o trabalho foi, portanto, retirada do banco de dados *Protein Data Bank* (Bernstein et al., 1977) utilizando o código PDB 1GRI (Sakai et al., 2002), onde encontram-se as informações referentes às coordenadas espaciais de cada átomo com a proteína em sua forma dimérica, viabilizando uma análise completa das possíveis interações, seja nos domínios já relatados na literatura ou em novas cavidades formadas apenas no dímero.

2.4.2 Docking Molecular (UCSF DOCK 6)

Para o cálculo de *docking* molecular, o programa utilizado foi o *UCSF Dock v 6.7* da Universidade de São Francisco – Califórnia e seus programas acessórios (Lang et al., 2009). Sua primeira abordagem simplista de campo de forças, com complementos sofisticados de cálculos complexos, como a Energia Livre de interação pelo método MM/GBSA foram decisivos para a escolha do programa, assim foi possível obter

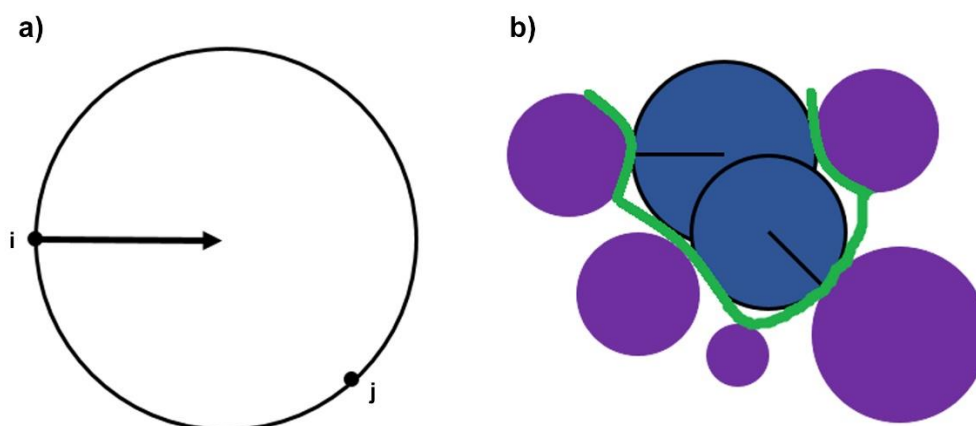
³ PDB das estruturas analisadas: 1BM2, 1BMB, 1CJ1, 1JYQ, 1TZE, 3MXC, 3N84, 3OVE, 1GCQ, 1IO6, 2VVK

resultados de excelente qualidade, melhores que alguns programas popularmente utilizados. Sua praticidade nesses cálculos permitem uma ampla amostragem de conformações, podendo ser implementado fundamentações estatísticas para o refinamento de resultados.

O DMS (*Display Midas System*) foi utilizado para o cálculo da superfície molecular do receptor com base no algoritmo desenvolvido por Richards e adaptado por Connolly (Shoichet, Bodian, & Kuntztz, 1992), onde uma esfera do tamanho de uma molécula de água com raio de 1.4 Å é usada como referência para mapear a superfície de Van der Waal da macromolécula. Além disso, o vetor normal de superfície em cada ponto foi calculado.

O programa SPHGEN (Kuntz et al., 1982) foi utilizado para a criação do conjunto de esferas (clusters) com base na superfície molecular gerada pelo DMS como mostrado na Figura 22, o tamanho das esferas foi calculado de acordo com o tamanho do vetor normal da superfície molecular e foram utilizadas nas próximas etapas como ponto de ancoragem do ligante e centro da caixa para o cálculo do GRID.

Figura 22 - SPHGEN, como calcular esferas. (a). Cada esfera é gerada tangente aos pontos de superfície i, j com o centro na superfície normal do ponto i. (b) Representação esquemática de um pequeno sítio de ligação formado por cinco átomos (roxo). As esferas em azul são geradas usando pontos da superfície molecular (verde) com seus centros que se encontram ao longo dos vetores normais da superfície.



Fonte: adaptado de dock.compbio.ucsf.edu/DOCK_6

As esferas foram calculadas sobre toda a superfície molecular, produzindo aproximadamente uma esfera por ponto vetor normal. As esferas podem estar contidas externamente a superfície quando em cavidades do lado interior das proteínas, desde que não tenha impedimentos estéricos. O Raio máximo de cada esfera foi de 4.0 Å e o mínimo 1.4 Å. Esta representação é, então, filtrada para manter apenas a maior esfera associada a cada átomo. O conjunto filtrado é agrupado e cada cluster resultante representa uma possível cavidade de interação no receptor.

O programa *Showbox* foi usado para definir a localização e o tamanho da caixa onde foi calculado o GRID. A caixa foi construída automaticamente centralizada de acordo com as cluster contendo as esferas selecionadas e nos permitiu a inclusão de margens extras em cada direção, para esse trabalho, foram adicionados 5Å em cada coordenada, tamanho suficiente para o ligante utilizado. O programa GRID é utilizado para gerar os arquivos de grade com informações necessárias para a primeira avaliação do score. A caixa construída através de *Showbox* foi então dividida em grades de 0.3Å cada uma contendo informações acerca de distância dos resíduos, necessário para evitar colisões e sobreposições de átomos do receptor com o ligante e as energia coulombianas com informações de cargas e distâncias dos resíduos próximos (Shoichet et al., 1992). A função de pontuação utilizada para esses cálculos é denominada *Grid Score* e baseia-se em parâmetros físicos, determinados a partir de um campo de força simplificado que utiliza apenas o termo de interação de Van der Waals combinado com a componente eletrostática como mostrado na equação 5.

$$E = \sum_{i=1}^{ligante} \sum_{j=1}^{receptor} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{D r_{ij}} \right) \quad (\text{eq. 5})$$

Cada termo é uma dupla somatória sobre os átomos i do ligante e os átomos j do receptor. A_{ij} e B_{ij} são parâmetros da repulsão e da atração de Van der Waals, r_{ij} é a distância entre os átomos i e j, q_i e q_j são as cargas pontuais dos átomos i e j, D é a função dielétrica e 332 é o fator que converte a energia eletrostática em quilocalorias por mol.

A próxima etapa de “encaixe” foi conduzida pelo algoritmo de *Anchor and Grow* (Ewing, Makino, Skillman, & Kuntz, 2001). Inicialmente, o ligante foi subdividido em fragmentos rígidos compostos por resíduos em que as ligações adjacentes não possuem

graus de rotatividade, e fragmentos flexíveis que ligam as estruturas rígidas e apresentam graus de rotatividade. Vale ressaltar que esse esquema de classificação é uma aproximação de primeira ordem da flexibilidade molecular, uma vez que alguma quantidade de flexibilidade pode existir em anéis não aromáticos. O maior fragmento rígido foi selecionado como “âncora” e utilizado para iniciar o encaixe do ligante no receptor, os demais fragmentos foram então “crescidos” e seus *scores* calculados, como mostrado na Figura 23.

Devido a simplicidade do primeiro cálculo, e ao alto número de possibilidades (número esse que depende do sítio e do ligante utilizado), esse processo foi repetido 2500 vezes, a fim de obter uma boa amostragem conformacional, podendo selecionar as melhores conformações por meio de uma abordagem estatística.

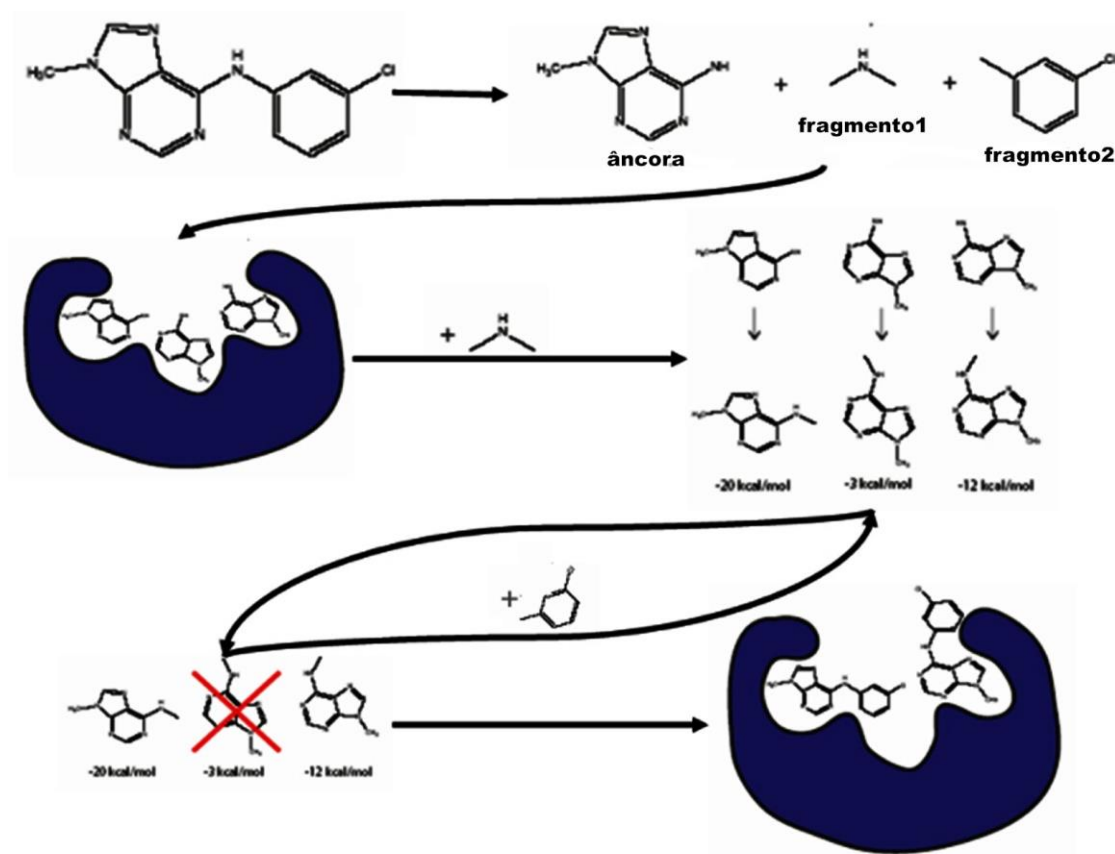
Adicionalmente foi utilizada a função *Amber Score* (Graves et al., 2008), que permite utilizar as configurações obtidas na primeira fase do processo e reclassificá-las, utilizando uma dinâmica molecular empregando o campo de força do AMBER. Neste processo os átomos distantes do local da ligação podem ser mantidos “congelados”, diminuindo o custo computacional no cálculo das energias desses átomos.

As principais vantagens da função *Amber Score* é que tanto o ligante quanto o sítio ativo do receptor podem ser flexíveis, permitindo que pequenos rearranjos estruturais reproduzam o chamado “ajuste induzido” durante a realização da pontuação assim como a possibilidade do cálculo da energia livre de interação utilizando o método MM/GBSA (Graves et al., 2008), que apesar das aproximações, pode ser utilizada como parâmetro para comparações teóricas-experimentais. A energia de interação do *Amber Score* foi calculada através da Equação 6 e será parâmetro para escolha, junto com os parâmetros experimentais, das melhores conformações em cada bolso de interação. Cada termo de energia é obtido através do campo de força do *Amber*, incluindo termos de solvatação utilizando modelos contínuos de solvatação⁴ (Graves et al., 2008).

$$E_{interação} = E_{complexo} - (E_{ligante} + E_{receptor}) \quad (\text{Eq. 6})$$

⁴ Maiores informações sobre a função *Amber Score* e cada termo de energia pode ser encontrado em (http://dock.compbio.ucsf.edu/DOCK_6/tutorials/amber_score/amber_score.htm)

Figura 23 - Algoritmo *Anchor and Grow*. Primeiro, a maior subestrutura rígida do ligante (a âncora) é identificada e em seguida suas camadas flexíveis. Na segunda etapa, a âncora é rigidamente orientada no sítio e por fim, no estágio de crescimento, as camadas flexíveis e as subunidades rígidas menores do ligante são construídas sobre as melhores orientações em relação ao receptor. Após o crescimento completo, a estrutura com menor score é selecionada como o melhor encaixe.



Fonte: adaptado de (dock.compbio.ucsf.edu)

2.4.3 Dinâmica Molecular do complexo Grb2 e ligantes.

Tendo em vista que *dockings* moleculares não contabilizam junto aos seus cálculos toda a flexibilidade do receptor nem as interações com solventes, possíveis apenas com o uso de solventes explícitos, é importante uma etapa de dinâmica molecular, verificando a validade do *docking* e proporcionando um ambiente onde as moléculas podem se arranjar conformacionalmente dentro dos bolsões.

Todas as etapas da dinâmica molecular, desde a preparação do sistema até a simulação foram realizadas com o pacote de programas GROMACS, versão 4.5.7 (Pronk et al., 2013) com os recursos computacionais do Grid Unesp. A necessidade de um campo de força e de um programa para parametrização dos ligantes nos conduziu à escolha do CHARMM27 (MacKerell et al., 1998) incluído no pacote GROMACS, e *Swissparam* (MacKerell et al., 1998), disponível através do *Web Server* do grupo de modelagem molecular SIB disponível em (<http://www.swissparam.ch/>). A otimização específica do programa de parametrização do *Swissparam* para o campo de força CHARMM foi o principal ponto para a sua escolha.

Os parâmetros de carga escolhidos para os resíduos foram padrões (representação da carga mais provável quando o resíduo se encontra em pH fisiológico). Todas as simulações foram realizadas em caixas cúbicas utilizando o modelo de água SPC Flexível e condições periódicas de contorno. Os sistemas foram neutralizados pela substituição de algumas moléculas de água por íons de cloro (Cl^-) e sódio (Na^+) que ao todo ficaram em uma concentração de 150 mM. Em geral, os sistemas foram compostos de aproximadamente 26.000 átomos.

Antes da dinâmica molecular, foram realizados 50000 passos de minimização de energia pelo método *steepest-descent* com todos os átomos flexíveis e 5000 passos de minimização de energia pelo método dos gradientes conjugados. Após os processos de minimização de energias, foram realizadas duas etapas de equilibração, uma MD de 100 ps com restrição de posições para os átomos da proteína e do ligante, de modo que o solvente ficasse livre para se ajustar de maneira correta, e uma outra MD de 100 ps sem restrições de posições para que o sistema se adaptasse às condições termodinâmicas impostas (relaxação).

As simulações de MD foram de 20 ns, tempo o suficiente para verificação das interações entre Grb2 e os ligantes Cumarina e Morina, uma vez que os resultados de RMN indicam que há uma dinâmica de encaixe e desencaixe do ligante devido à sua interação moderada permitindo o cálculo de saturação via STD. Foram utilizados passos de integração de 2 fs, com temperatura e pressão mantidas constantes, respectivamente, utilizando *v-rescale* e Parrinello-Rahman com os valores de 295K e 1 atm, as coordenadas e as velocidades foram salvas a cada 10 ps e as interações eletrostáticas de longo alcance foram tratadas com o algoritmo *PME-Switch*.

2.5 Resultados e discussões.

De acordo com a metodologia aplicada neste trabalho, os experimentos *in silico* desenvolvidos foram fundamentados em resultados *in vitro*, os quais caracterizaram a interação da proteína Grb2 em sua forma dimérica com os ligantes Cumarina e Morina, mas não dizem respeito ao domínio ou sítio de interação. Diante desse fator, para uma total coerência de resultados, a estrutura utilizada deve conter a proteína em sua forma dimérica.

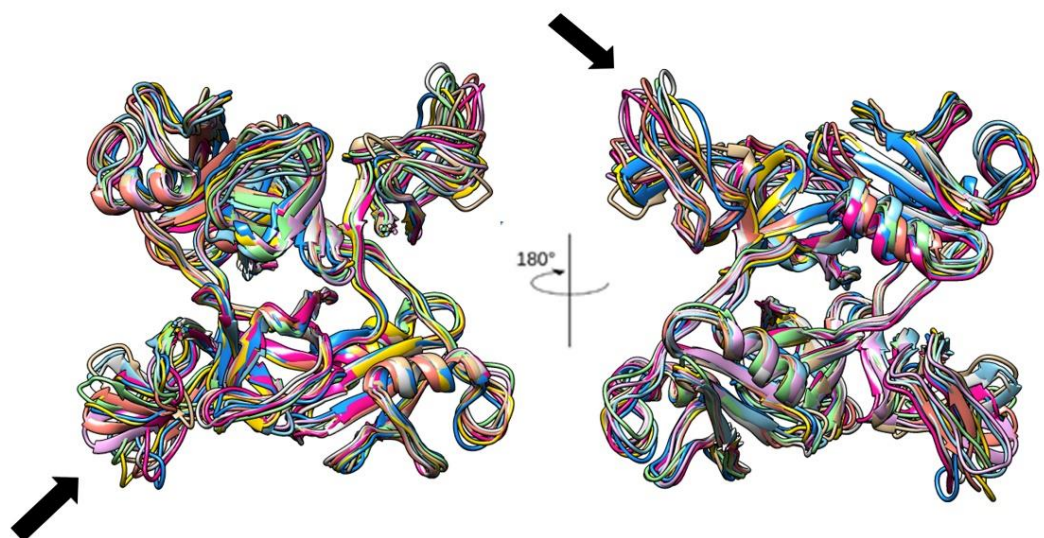
No entanto, é de conhecimento geral que proteínas não são macromoléculas rígidas. Suas estruturas em solução são flexíveis, podendo ter flutuações em suas cadeias laterais e pequenas mudanças conformacionais em suas estruturas secundárias e terciárias, acessando diferentes conformações quando em solução. Outros fatores como ligantes e alteração do meio químico podem também contribuir para a ocorrência de mudanças conformacionais.

O passo inicial foi, portanto, fazer uma dinâmica molecular, permitindo a hidratação e relaxação da proteína, rearranjando seus resíduos e podendo expor cavidades antes não vistas em sua estrutura cristalográfica, fornecendo informações inclusive sobre flutuações de domínios como visto na Figura 24.

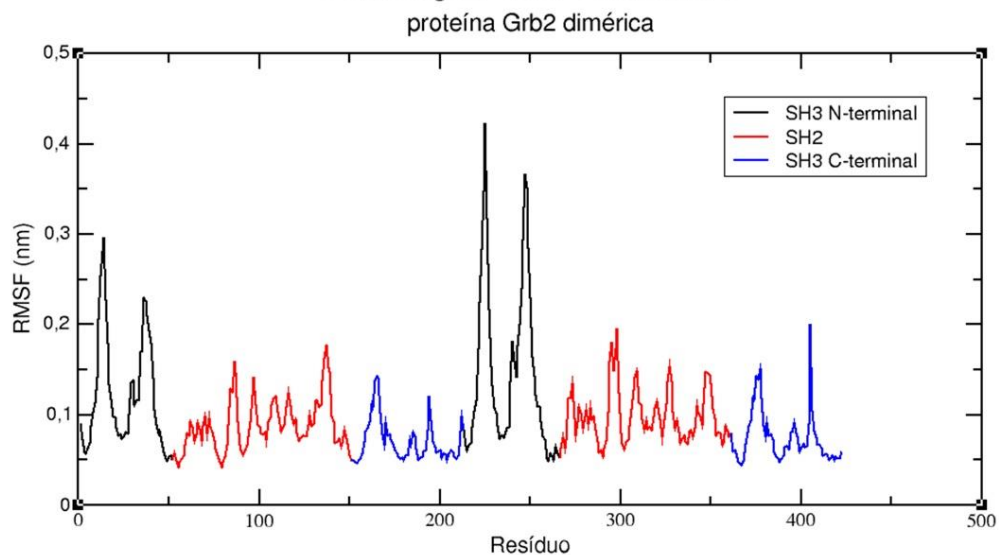
Uma verificação mais detalhada dos resultados de dinâmica molecular mostra que as estruturas de Grb2 podem ser clusterizadas conforme a distribuição de suas conformações. Ao todo, a trajetória foi dividida em 89 clusters, sendo destacado na Figura 25 os 3 clusters com maior número de acessos.

O critério de seleção da estrutura utilizada nas próximas etapas foi então a acessibilidade da estrutura em solução durante a dinâmica molecular. Toda essa primeira avaliação se fez necessária por ser um experimento de interação *in silico* inédito, sem um parâmetro inicial na literatura com ligantes de *scaffold* parecidos com Cumarina e Morina ou até mesmo interações com a proteína Grb2 em sua forma dimérica.

Figura 24 - Análise estrutural da proteína GRB2 em solução por dinâmica molecular. (A) Sobreposição de diferentes estruturas retiradas da dinâmica molecular utilizando o PDBid 1GRI sem ligantes. O domínio SH3 N-terminal destacado pelas flechas mostra-se com um maior grau de flexibilidade quando comparado aos outros resíduos. SH2 e SH3 C-terminal são domínios com maior estabilidade em solução, com poucas variações estruturais. (B) RMSF da trajetória, obtida utilizando a ferramenta *g_rmsf* do pacote GROMACS (Pronk et al., 2013). Neste fica evidenciado o alto grau de flexibilidade do domínio SH3 N-terminal.

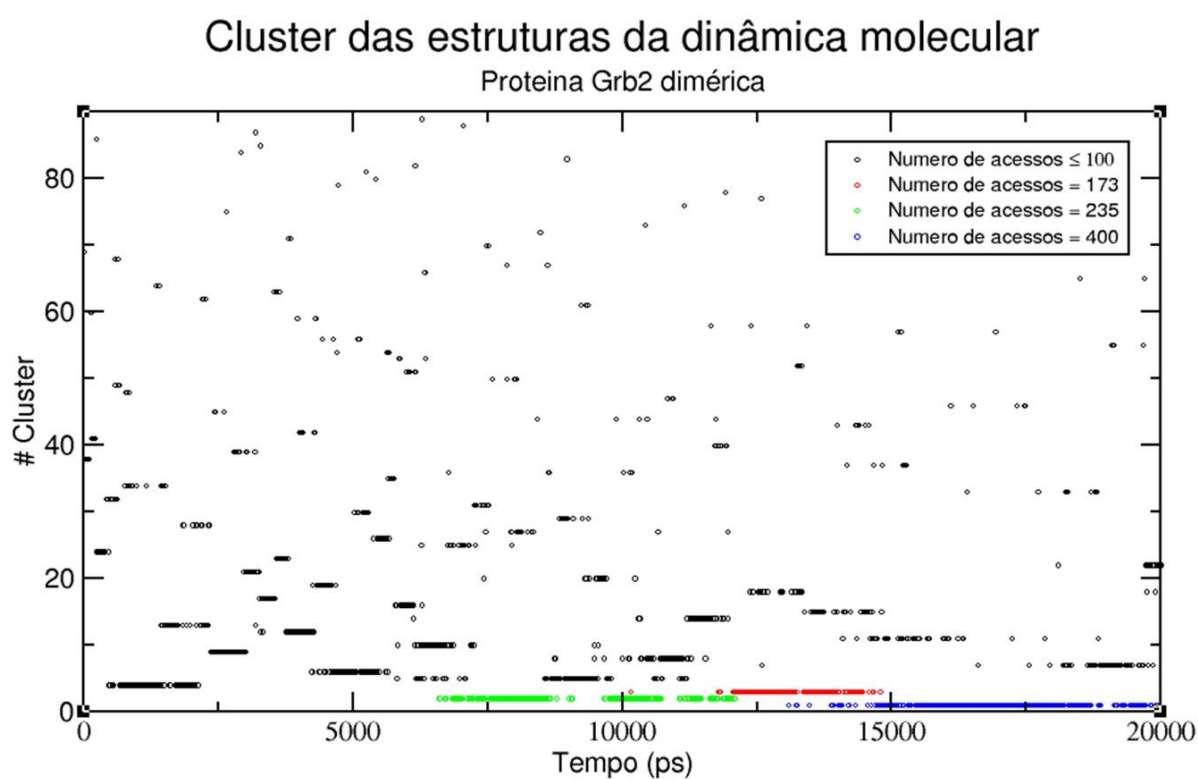


Flutuação dos Resíduos



Fonte: Elaborada pelo autor

Figura 25 - Clusterização das conformações estruturais de Grb2. A dinâmica molecular da proteína Grb2 é dividida em 89 clusters com critério de separação baseados em sua estrutura conformacional. Por cores, estão separados os 3 clusters com um maior número de acessos de uma mesma estrutura, calculados utilizando g_cluster do pacote GROMACS (Pronk et al., 2013), sendo em azul o maior com um total de 400 acessos, em verde a segunda estrutura mais acessada com um total de 235, em vermelho o terceiro maior com 173 e os demais são iguais ou menores que 100 acessos. A simulação computacional teve um tempo de 20000 ps.



Fonte: Elaborada pelo autor

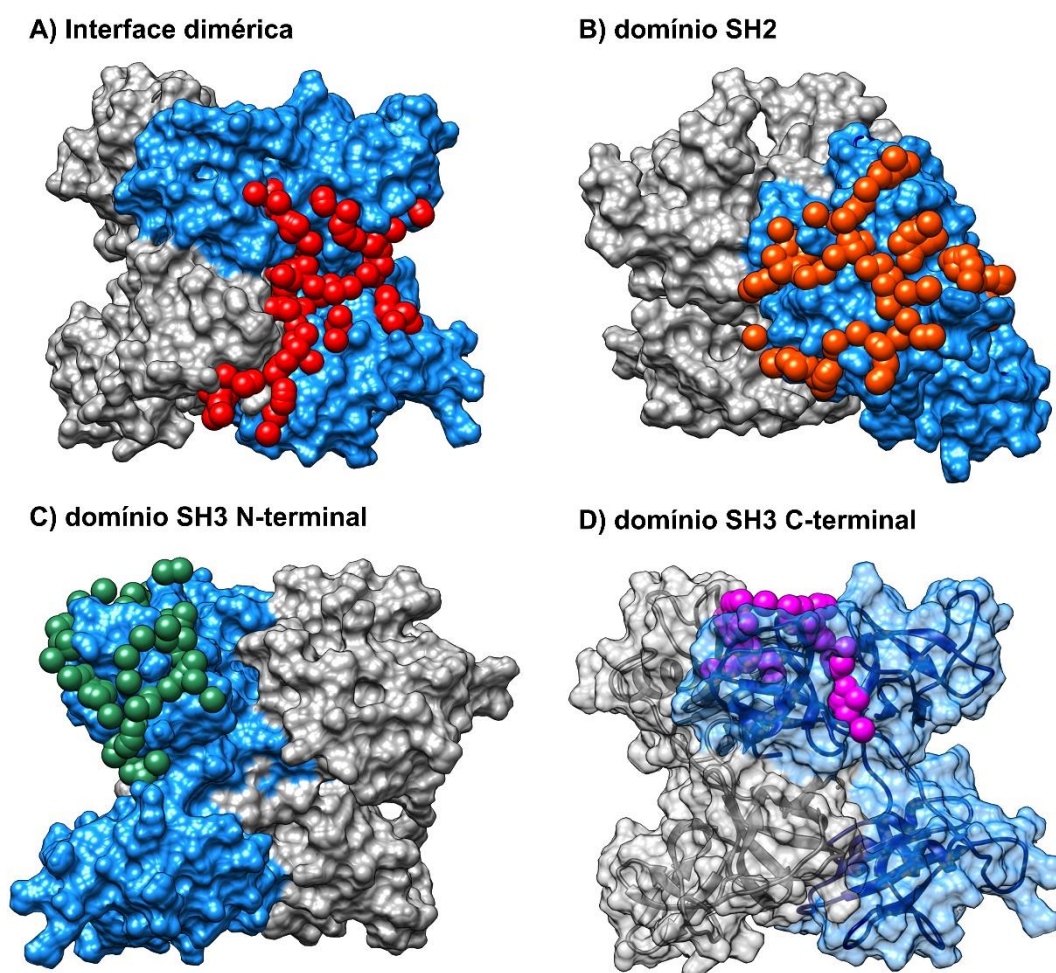
2.5.1 Identificação de cavidades nos domínios para o docking molecular

A estrutura tridimensional hidratada e relaxada permite a busca por possíveis cavidades para o processo de *docking* molecular, o que foi realizado pelo programa SPHGEN. Ao todo, foram geradas 39 diferentes regiões, nomeadas por bolsões de interação, espalhados por toda superfície molecular da proteína. Uma análise inicial mostrou que destes 39 bolsões, apenas 9 eram fisicamente aceitáveis, uma vez que os parâmetros investigados foram a capacidade de receber o ligante, dado o tamanho do cluster, e esferas definidas para a ancoragem.

Ainda, dentre esses 9 bolsões fisicamente aceitáveis, os 4 maiores e principais contemplam regiões específicas dos domínios monoméricos, isto é, os domínios SH2 e SH3s isolados, assim como as regiões de interface proteína/proteína, como mostrado na Figura 26. Os outros 5 são repetições de clusters em regiões idênticas e espelhadas devido ao efeito da dimerização da proteína Grb2. Portanto os 4 maiores bolsões foram selecionados e utilizados para os processos de *docking* de dinâmica molecular.

Na Figura 26, as esferas vermelhas estão localizadas no bolsão formado pela interface de ligação monômero-monômero, evidenciando a importância da escolha do PDB 1GRI, que contempla o dímero da proteína GRB2. As esferas rosas abrangem a cavidade formada no domínio SH3 C-terminal, e em verde são representadas as esferas que ocupam as cavidades do domínio SH3 N-terminal, bolsão esse que, apesar de largo, não possui uma grande profundidade com bolsões bem pronunciados. No domínio SH2 se encontra o maior cluster, representado pelas esferas em laranja.

Figura 26 - Possíveis cavidades para o *docking* do ligante. A busca por cavidade foi realizada por meio do programa SPHGEN (Kuntz et al., 1982), sendo as maiores cavidades encontradas na estrutura da proteína GBR2 apresentadas por esferas. Esferas de mesma cor compõem um mesmo bolsão, constituindo uma única região onde foi realizado o docking molecular. Elas também funcionam como pontos de ancoragem para o ligante na macromolécula durante o processo de “*anchor-and-grow*”. Para facilitar a visualização, os bolsões assumem cores referentes à região da proteína, sendo em vermelho uma região da interface monômero-monômero, laranja o domínio SH2, verde o domínio SH3 N-terminal e em rosa o domínio SH3 C-terminal. As superfícies foram coloridas por seus monômeros, sendo o monômero A colorida em azul e o monômero B colorido em cinza. Foi adicionado também a transparência na superfície em D devido ao fato de que algumas esferas rosas se encontram internamente a superfície molecular.



Fonte: Elaborada pelo autor. PDBid 1GRI.

As possíveis interações foram calculadas utilizando o programa DOCK6, utilizando um potencial simplificado (Eq. 4) para o cálculo do *GridScore* e repetido 2500 vezes, sendo avaliadas 5.000 configurações em cada corrida de *docking*. Para cada cluster foram obtidas um total de 2500 conformações, totalizando 10000 conformações separadas em 4 clusters para cada molécula.

Para um refinamento dos resultados, o campo de força *Amber Score*, incluído no pacote DOCK6 foi utilizado para o cálculo das Energias de Interação utilizando o método MM/GBSA (Graves et al., 2008). Os resultados para cada bolsão de interação são ordenados considerando suas energias de interação e frequência de acesso as estruturas durante o docking molecular.

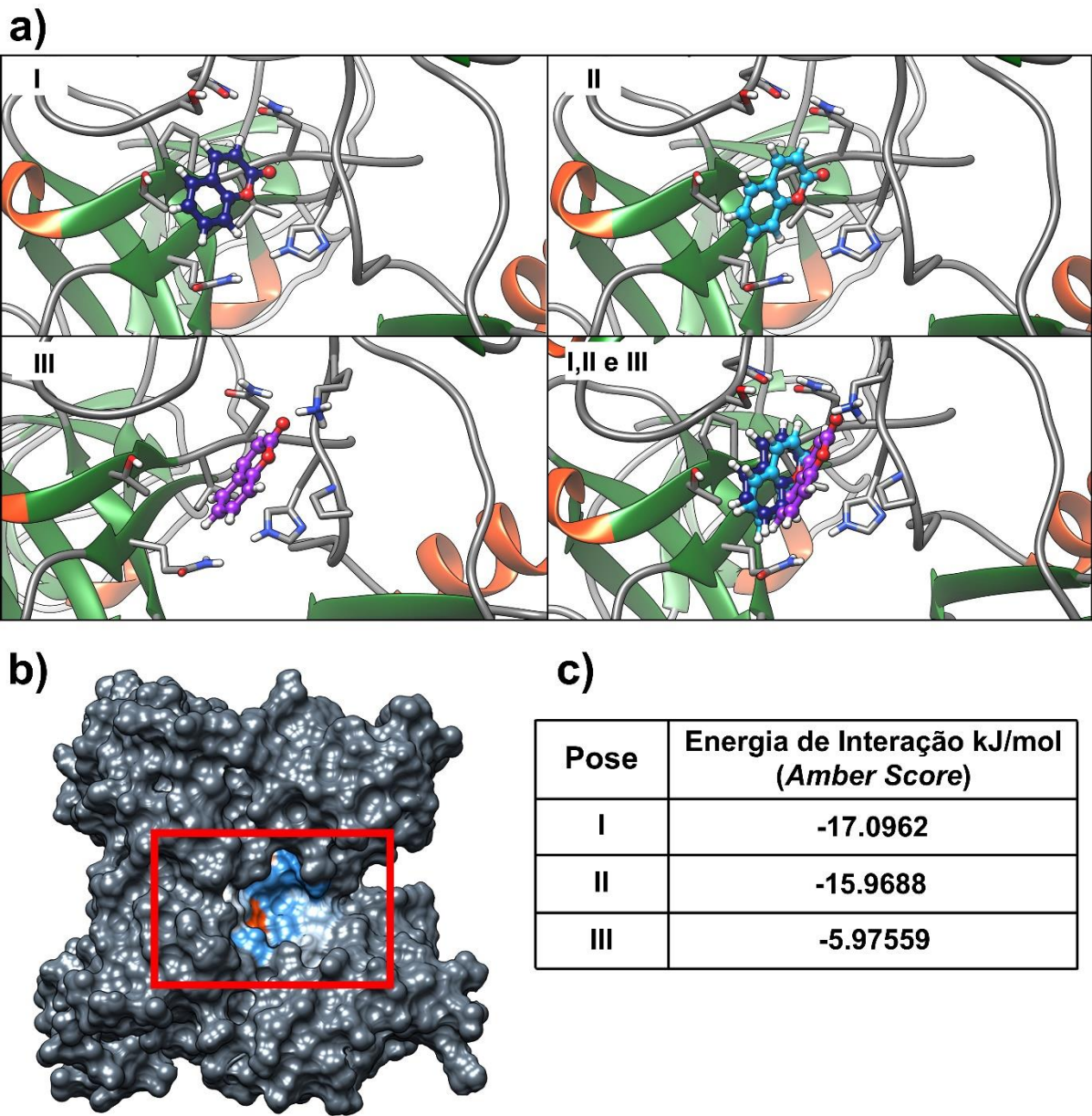
Para a região de interface monômero-monômero, três diferentes disposições estruturas das moléculas de Cumarina e Morina foram encontradas, como mostrado nas Figuras 27 e 28 respectivamente. Nessa região, apesar da presença de alguns resíduos de caráter hidrofóbicos, predomina-se os resíduos de caráter hidrofílicos expostos ao solvente, como mostrados em 27b e 28b a partir da superfície molecular.

Próximo a essa cavidade, há um resíduo de triptofano (W278) com distâncias médias para as 3 diferentes poses da molécula de Cumarina de 8Å e da molécula de Morina de 4.5Å. Esses valores indicam que o parâmetro utilizado como resíduo de Triptofano corrobora com essa cavidade.

As disposições estruturais das três diferentes poses de ambas as moléculas também não são como as esperadas pelos resultados de STD-RMN (Figura 19), onde é sugerido um encaixe do ligante em cavidades que os envolvem-se por completo seus anéis, ficando com a ramificação para fora do bolso no caso da estrutura da Morina.

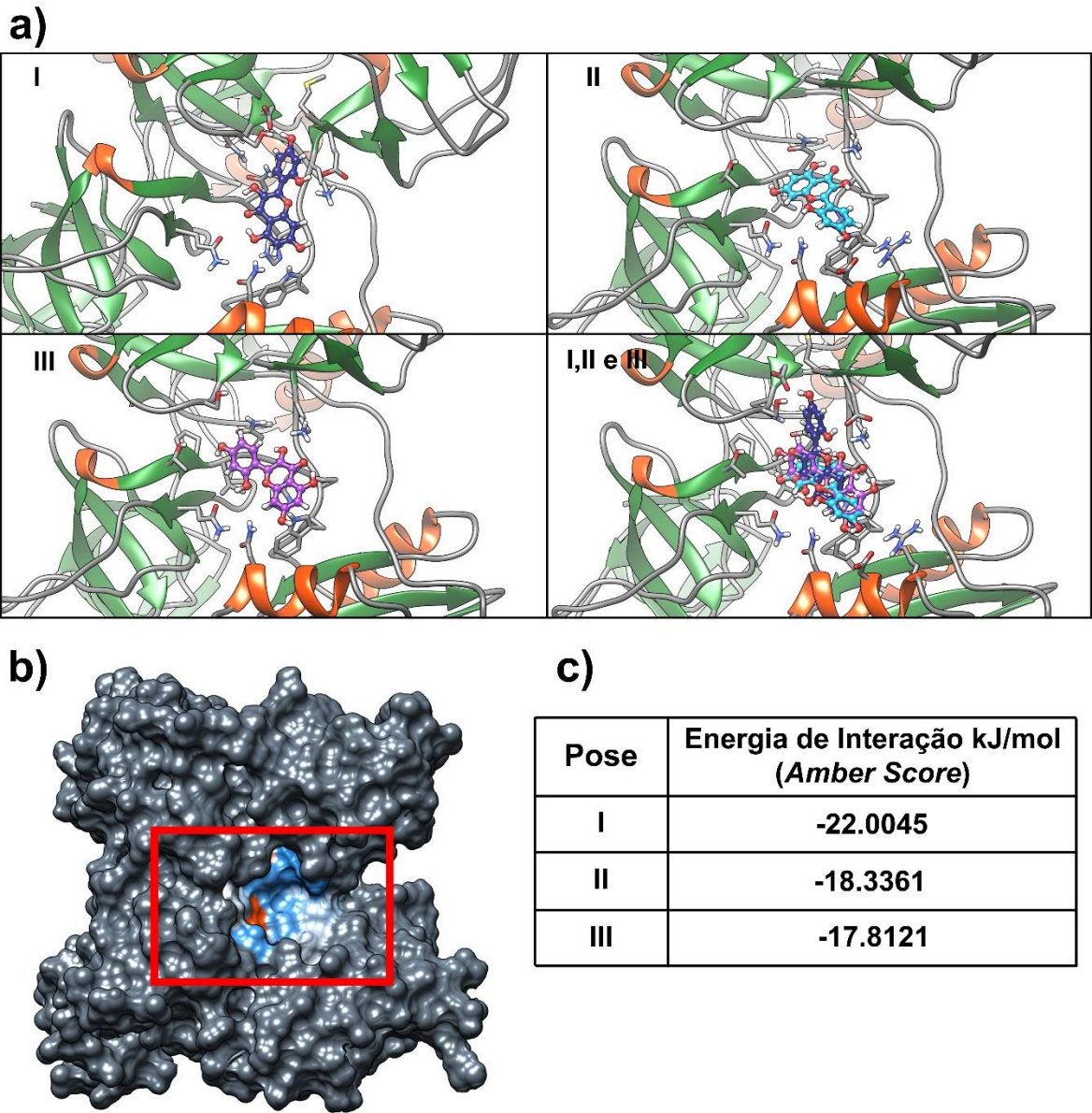
Esses resultados mostram então que, os melhores resultados de docking para a bolsão contido região dimérica de Grb2, calculados utilizando o parâmetro computacional Energia de interação, não corroboram com todos os parâmetros experimentais obtidos por RMN e fluorescência.

Figura 27 – Melhores conformações do *Docking* da molécula de Cumarina no bolsão contido na interface dimérica. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo *Amber Score* para cada pose de A.



Fonte: Elaborada pelo autor.

Figura 28 – Melhores conformações do *Docking* da molécula de Morina no bolsão contido na interface dimérica. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo *Amber Score* para cada pose de A.



Fonte: Elaborada pelo autor.

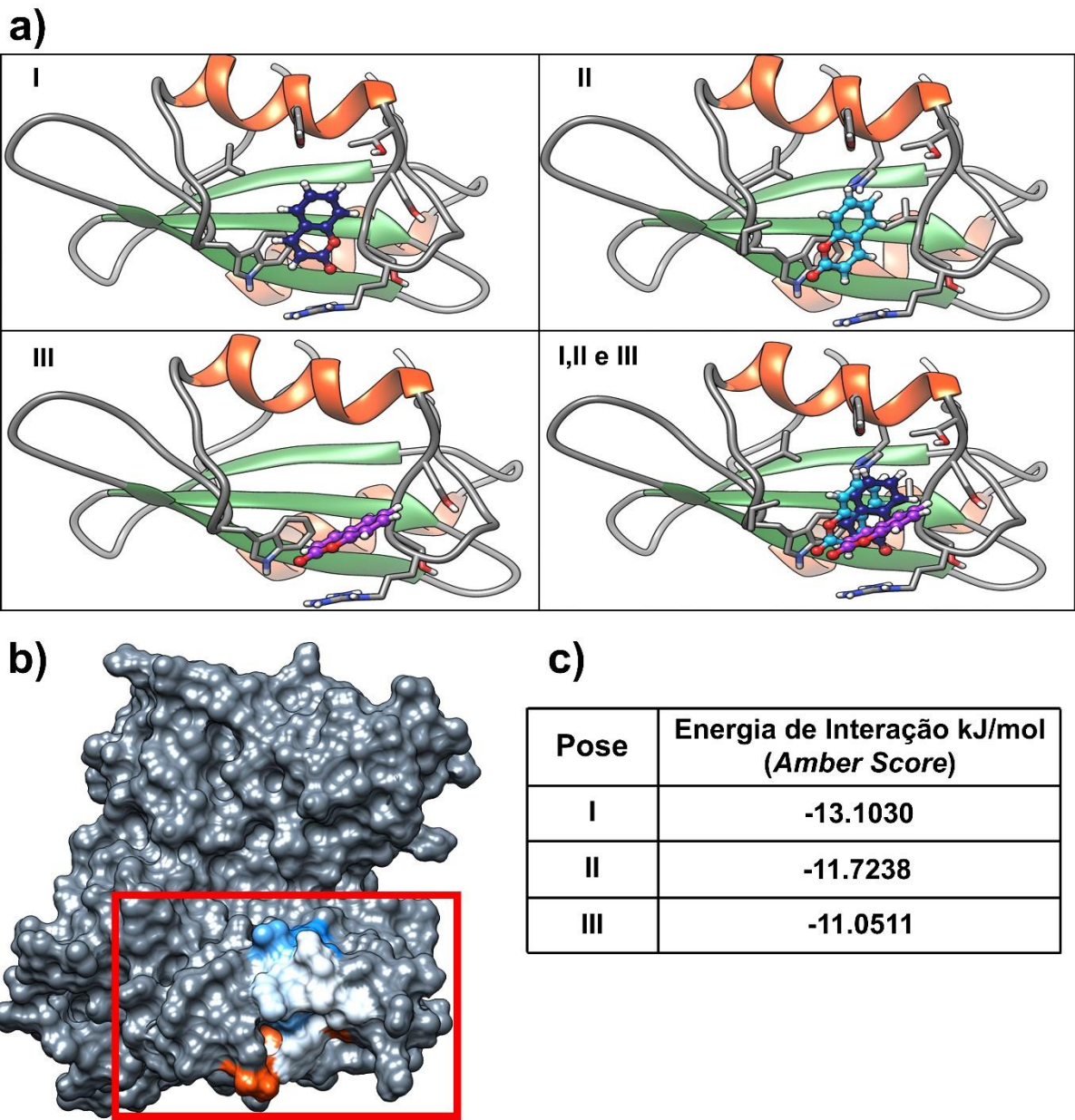
O bolsão contido no domínio SH2 de Grb2 apresenta o segundo melhor resultado de energia de interação computacional para ambas as moléculas. Suas três melhores poses obtidas por docking são mostradas nas Figuras 29 e 30.

A análise desse bolso mostrou que há predominância de resíduos hidrofóbicos, proporcionando um ambiente propício para receber interações com efeitos hidrofóbicos predominantes. Também está presente nesse bolso um triptofano (W121) com distâncias médias das três diferentes poses para a molécula de Cumarina de 3.3 Å e 3.8 Å para Morina. Essas distâncias caracterizam interações de Van der Waals (VdW) entre os ligantes e o resíduo de triptofano. A disposição dos anéis aromáticos de W121 paralela aos anéis do ligantes também caracteriza possíveis interações do tipo *stacking- π* . Uma vez que possivelmente ocorre essas interações, os resultados de docking concordam com o parâmetro experimental de que existia um triptofano no bolsão, o qual tinha seus movimentos alterados devido a presença do ligante.

Contudo, os resultados dos epítomos do RMN ainda não estão de total acordo com os resultados de docking. O encaixe do ligantes no bolsão calculados computacionalmente não são como os sugeridos experimentalmente.

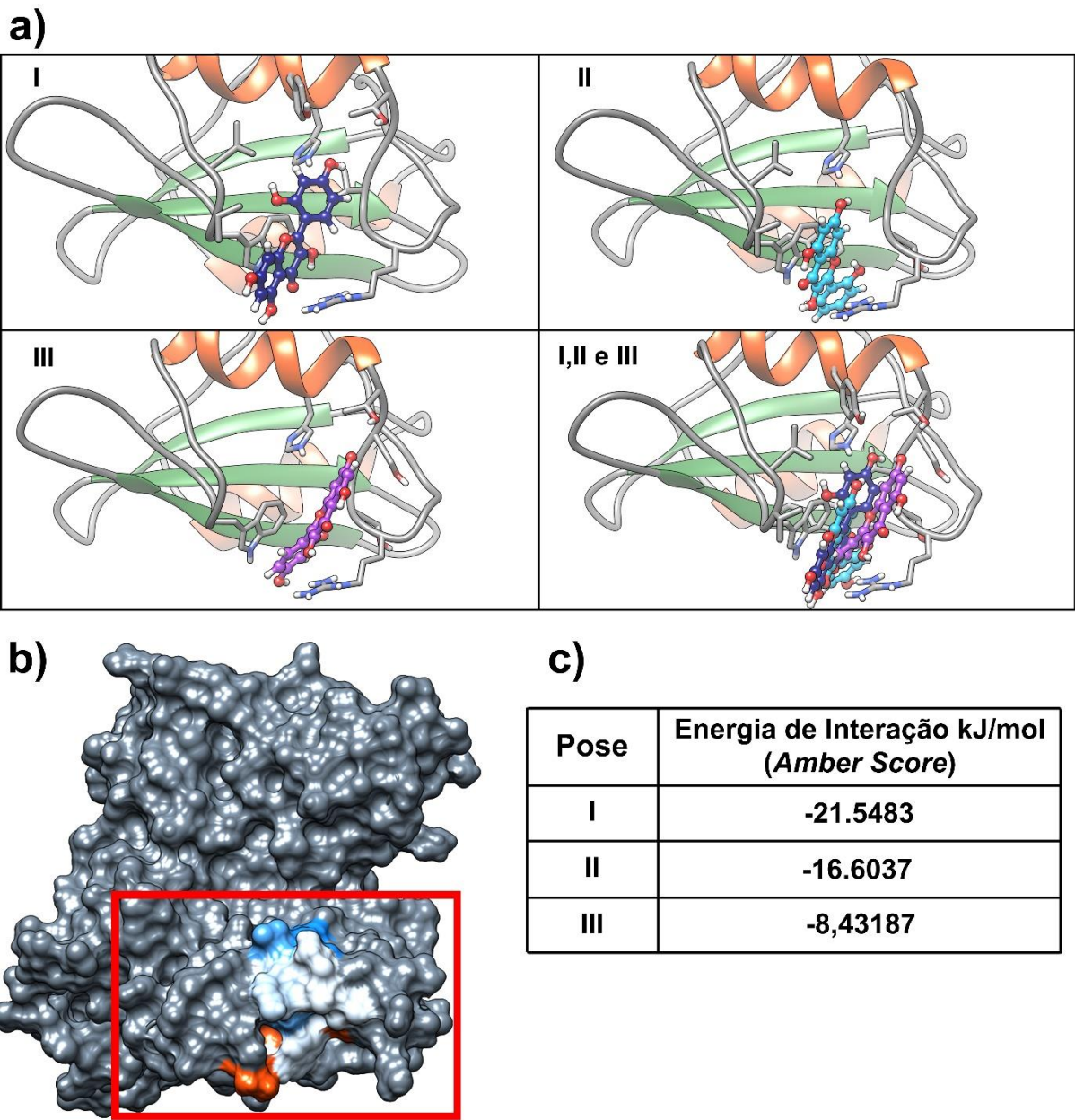
Ainda sobre o domínio SH2 da proteína GRB2, a literatura relata experimentos realizados para análises termodinâmicas e mutacionais da ligação do domínio SH2 de Grb2 (McNemar et al., 1997). Dentre os resultados encontrados, um deles mostra a hidrofobicidade do domínio SH2 de Grb2 e também cita o resíduo de Triptofano W121 localizado no coil que dá origem ao bolsão hidrofóbico, e que pode ser visualizado em nosso resultado de *docking*. É descrito ainda que esse resíduo não só preenche a bolsa como também direciona os ligantes para o sítio (Rahuel J, et al., 1996). Informações essas a serem confirmadas na próxima etapa, onde será realizadas simulações computacionais utilizando o resultado do docking molecular.

Figura 29 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH2. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo Amber Score para cada pose de A.



Fonte: Elaborada pelo autor.

Figura 30 – Melhores conformações do *Docking* da molécula de Morina no bolsão do domínio SH2. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo *Amber Score* para cada pose de A.



Fonte: Elaborada pelo autor.

O docking das ligantes no domínio SH3 N-terminal é mostrado nas Figuras 31 e 32. Em resultados anteriores (Figura 24), esse domínio se mostrou flexível, mudando conformacionalmente ao longo da dinâmica molecular. Entretanto, na estrutura utilizada para o docking molecular, existem bolsões com tamanhos suficientes para receber os ligantes.

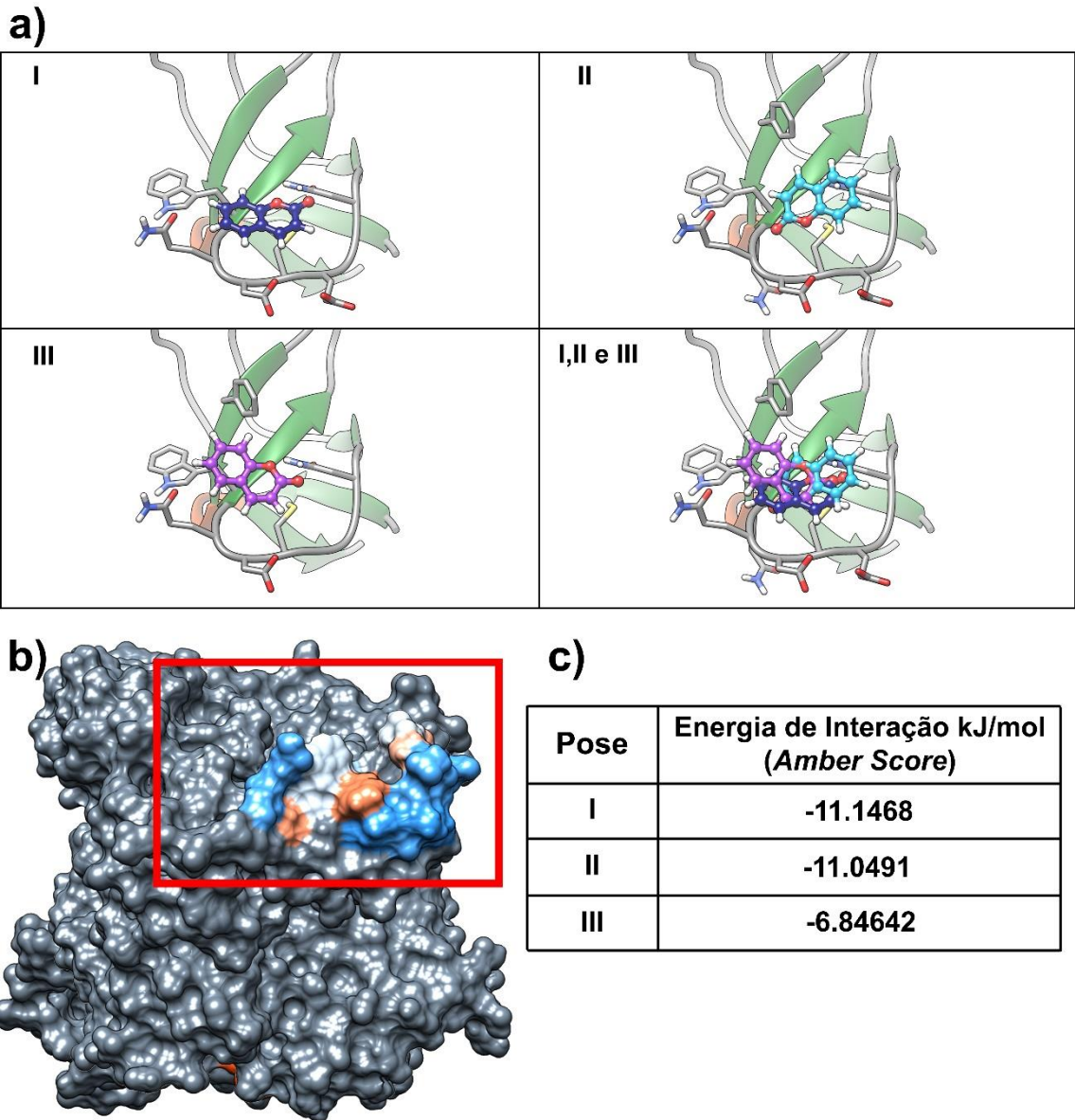
O resultado mostrou que essa região tem uma distribuição de resíduos hidrofóbicos e hidrofílicos por seus bolsões, fazendo com que haja regiões com características hidrofóbicas predominantes e regiões com características hidrofílicas predominantes em um mesmo bolsão.

Diferente dos outros dois *dockings* já relatados, as poses encontradas para o docking do ligante Morina em SH3 N-terminal mostra uma única corrida de 2500 testes, o encaixe em diferentes regiões, e que, apesar da mudança de região, as energias de interação calculadas são próximas.

As análises de interação mostram novamente um resíduo de triptofano (W36) nas proximidades do bolsão testado. A distância média desse resíduo para as três diferentes poses de Cumarina no bolso é de 7Å, para a pose I da Morina é de 20Å e para a pose II de 5Å. Esse resultado mostra que, diferente do que aconteceu com W121 do domínio SH2, W23 de SH3 N-terminal na conformação encontrada, não estaria interagindo com o ligante. Entretanto, o fato de estar contido no bolsão, não invalida o docking quando comparados com os parâmetros de fluorescência. Exceto a pose II da molécula de Morina, o qual mostra uma ancoragem em uma região do domínio em que não há resíduos de triptofanos, invalidando a conformação.

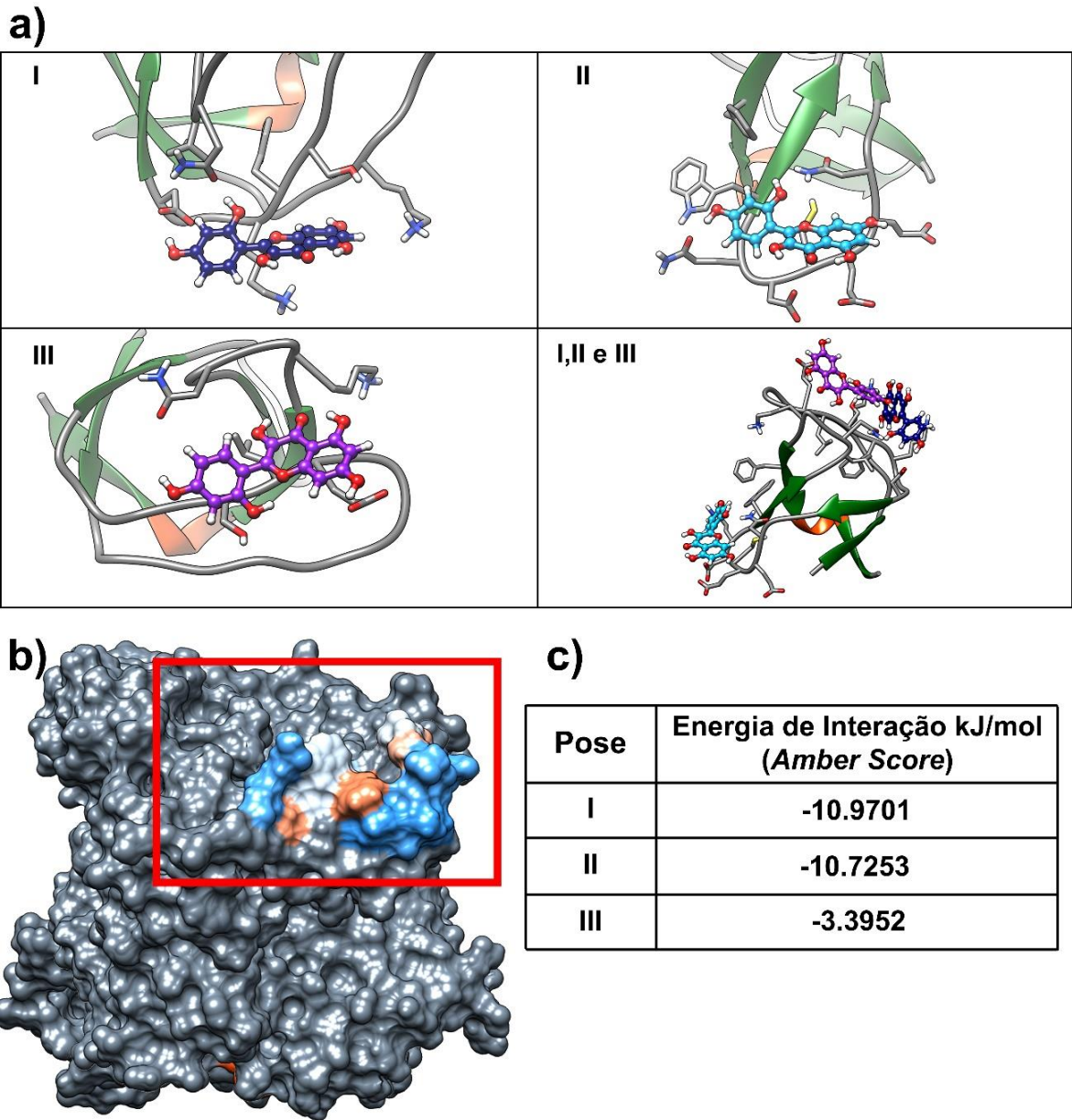
Os epítomos dos resultados experimentais de RMN-STD utilizados também como parâmetros para análise dos resultados, mais uma vez não corrobora com o docking molecular. Devido a SH2 ter bolsões poucos pronunciados e mais superficiais, ambas os ligantes não encaixaram nas cavidades conforme o sugerido pelos experimentos.

Figura 31 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH3 N-terminal. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo *Amber Score* para cada pose de A.



Fonte: Elaborada pelo autor.

Figura 32 – Melhores conformações do Docking da molécula de Cumarina no bolsão do domínio SH3 N-terminal. (A) A proteína Grb2 é colorida por suas estruturas secundárias, sendo verde as folhas- β , laranja as α -hélices e coils em cinza. 2500 testes de docking foi realizado e as 3 poses com menores valores de Energia de Interação e maiores frequências de acesso são mostradas. (B) Superfície molecular de grb2 com destaque na região onde o ligante foi ancorado. As cores em destaque representam a hidrofobicidade dos resíduos indo do azul para resíduos hidrofílicos para o laranja representando os resíduos hidrofóbicos. (C) Valores das energias de interação calculadas pelo *Amber Score* para cada pose de A.



Fonte: Elaborada pelo autor.

O bolsão do domínio SH3 C-terminal, apesar de ter mostrado esferas grandes o suficiente para ancoragem dos ligantes, não teve sucesso na etapa de docking molecular. As poucas conformações obtidas computacionalmente apresentaram problemas estéricos com as outras regiões de Grb2.

Por fim, os resultados de docking mostraram diferentes conformações dos ligantes espalhadas pelos diferentes domínios. Ainda, cada domínio foi ranqueado utilizando como parâmetro a Energia de Interação obtida por métodos computacionais, resultando em três diferentes conformações para a molécula de Cumarina e quatro para a Morina. Quando utilizados os parâmetros experimentais, o esse número se reduz a 3 para Cumarina e 3 para Morina, pois em umas das poses encontradas no domínio SH3 N-terminal não havia um resíduo de triptofano próximo. Vale ressaltar ainda que, cada uma dessas conformações obtidas, não estão totalmente em acordo com os resultados experimentais, se fazendo necessário então uma próxima etapa de dinâmica molecular para verificar a dinâmica dos ligantes nos bolsões.

2.5.2 Dinâmica Molecular.

Com os resultados de *docking* apresentados na sessão anterior, a dinâmica molecular foi uma segunda etapa de avaliação de interações, que possibilitou pequenas mudanças conformacionais na estrutura de Grb2, verificando se a interação dos ligantes com Grb2 permanece estável, ou se, simplesmente o ligante se dissocia do bolsão de interação.

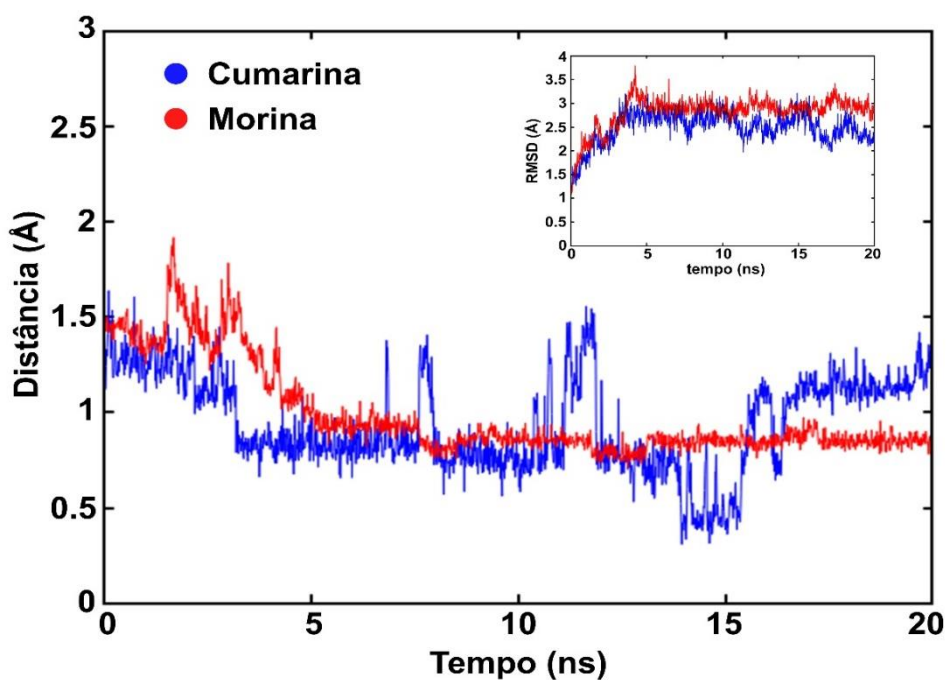
Os tempos de simulação, da ordem de ns, foram devido aos tempos de relaxamento captados pelos experimentos de RMN. Os tempos de interações e suas intensidades médias, fizeram com que a dinâmica de equilíbrio da interação do ligante seja a responsável pela captação de sinais. Vale ressaltar ainda que, os resultados mostrados a partir dessa sessão mostram apenas o domínio SH2 da proteína Grb2 para uma melhor visualização dos mecanismos e resíduos pertencente as interações.

Os resultados mostraram que, surpreendentemente, todos os complexos formados na etapa de *docking* permanecem por cerca de 10ns e são dissociados, com os ligantes sendo repelidos dos seus respectivos sítios. Salvo o docking de ambos os

ligantes no domínio SH2, que durante a dinâmica molecular, induziram mudanças conformacionais no domínio, expondo um bolso hidrofóbico, o qual ambos os ligantes se encaixaram.

Esse resultado é mostrado na Figura 33, em que a distância entre o centro de massa dos ligantes com o centro de massa da proteína Grb2 foi calculado, assim como o RMSD ao longo da simulação. A diminuição das distâncias e o aumento do RMSD mostraram que ambos os ligantes deixam o bolsão em que estavam encaixados e movem-se para essa nova cavidade.

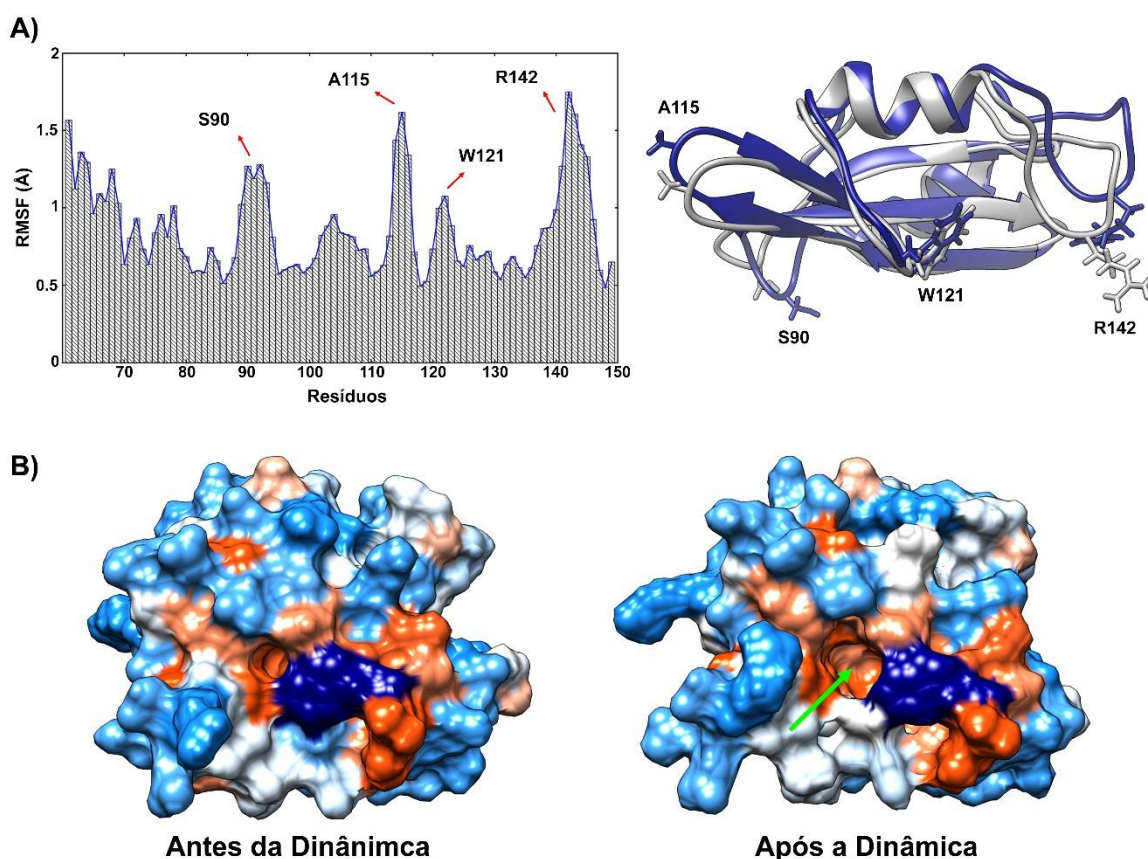
Figura 33 - Dinâmica molecular da interação dos ligantes com o domínio SH2. A distância entre o centro de massa dos ligantes Cumarina e Morina com o centro de massa do domínio SH2 da proteína Grb2 é calculado. Seus valores indicam uma aproximação do ligante com a proteína, mostrando que os ligantes se dissociam do sítio em que estava encaixado pelo docking molecular e se associam a um novo bolsão, induzido pela presença desses. O RMSD do sistema também é calculado para e sua variação também mostra que há pequenas mudanças conformacionais até a estabilização de seu novo valor. Os parâmetros calculados para o ligante Cumarina estão representados em azul e para a Morina em vermelho.



Fonte: Elaborada pelo autor

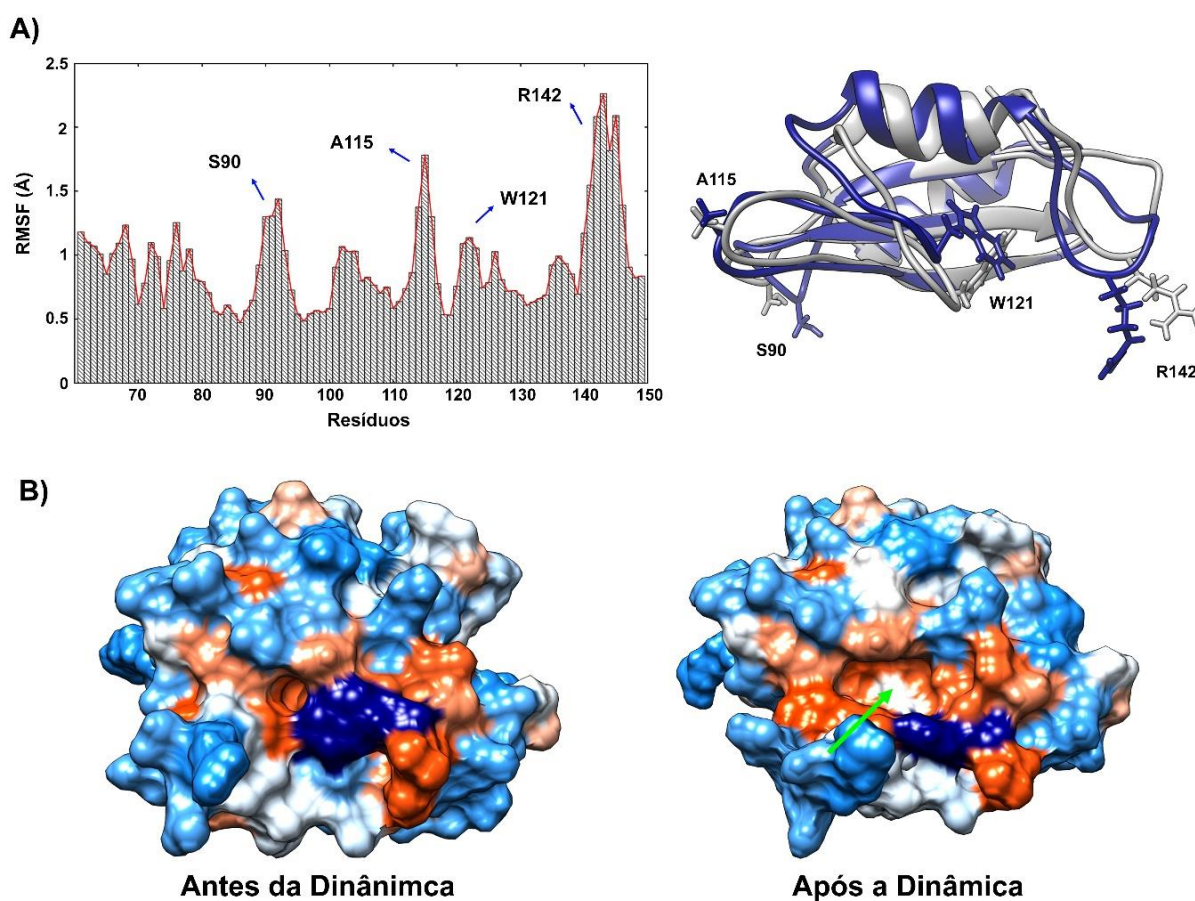
Para verificar as mudanças que os ligantes induziram no domínio SH2 da proteína Grb2 durante a dinâmica molecular, foi analisado a sobreposição de cadeias laterais e a flutuação dos resíduos, como mostrado nas Figura 34 e 35.

Figura 34 – Mudanças do domínio SH2 com a presença da Cumarina. (A) O perfil de flutuação dos resíduos durante a dinâmica molecular é calculado, mostrando os resíduos que mais sofreram alterações em suas cadeias laterais. Acompanhado do gráfico, a sobreposição das estruturas antes da dinâmica e após a dinâmica mostram exatamente onde estão localizados esses resíduos. Em cinza é representada a estrutura antes da dinâmica e em azul após a dinâmica. (B) Superfície molecular antes e após a dinâmica molecular mostrando que as mudanças conformacionais vistas em A alteram a superfície, gerando um novo bolso hidrofóbico. O Resíduo de triptofano (W121) é destacado em azul escuro. As superfícies foram coloridas de acordo com a hidrofobicidade dos resíduos, indo do laranja para resíduos hidrofóbicos para o azul representando os resíduos hidrofóbicos.



Fonte: Elaborada pelo autor

Figura 35 – Mudanças do domínio SH2 com a presença da Morina. (A) O perfil de flutuação dos resíduos durante a dinâmica molecular é calculado, mostrando os resíduos que mais sofreram alterações em suas cadeias laterais. Acompanhado do gráfico, a sobreposição das estruturas antes da dinâmica e após a dinâmica mostram exatamente onde estão localizados esses resíduos. Em cinza é representada a estrutura antes da dinâmica e em azul após a dinâmica. (B) Superfície molecular antes e após a dinâmica molecular mostrando que as mudanças conformacionais vistas em A alteram a superfície, gerando um novo bolso hidrofóbico. O Resíduo de triptofano (W121) é destacado em azul escuro. As superfícies foram coloridas de acordo com a hidrofobicidade dos resíduos, indo do laranja para resíduos hidrofóbicos para o azul representando os resíduos hidrofóbicos.



Fonte: Elaborada pelo autor

Os resultados mostraram que apesar de ser no mesmo sítio, causando mudança nos mesmos resíduos, a superfície molecular é alterada de modo diferente, adaptando-se aos respectivos ligantes. A Cumarina é responsável por induzir a abertura de um bolsão hidrofóbico grande o suficiente para envolvê-la. Já a Morina, induz mudanças que expõem o bolsão hidrofóbico, porém não grande o suficiente para se encaixar completamente.

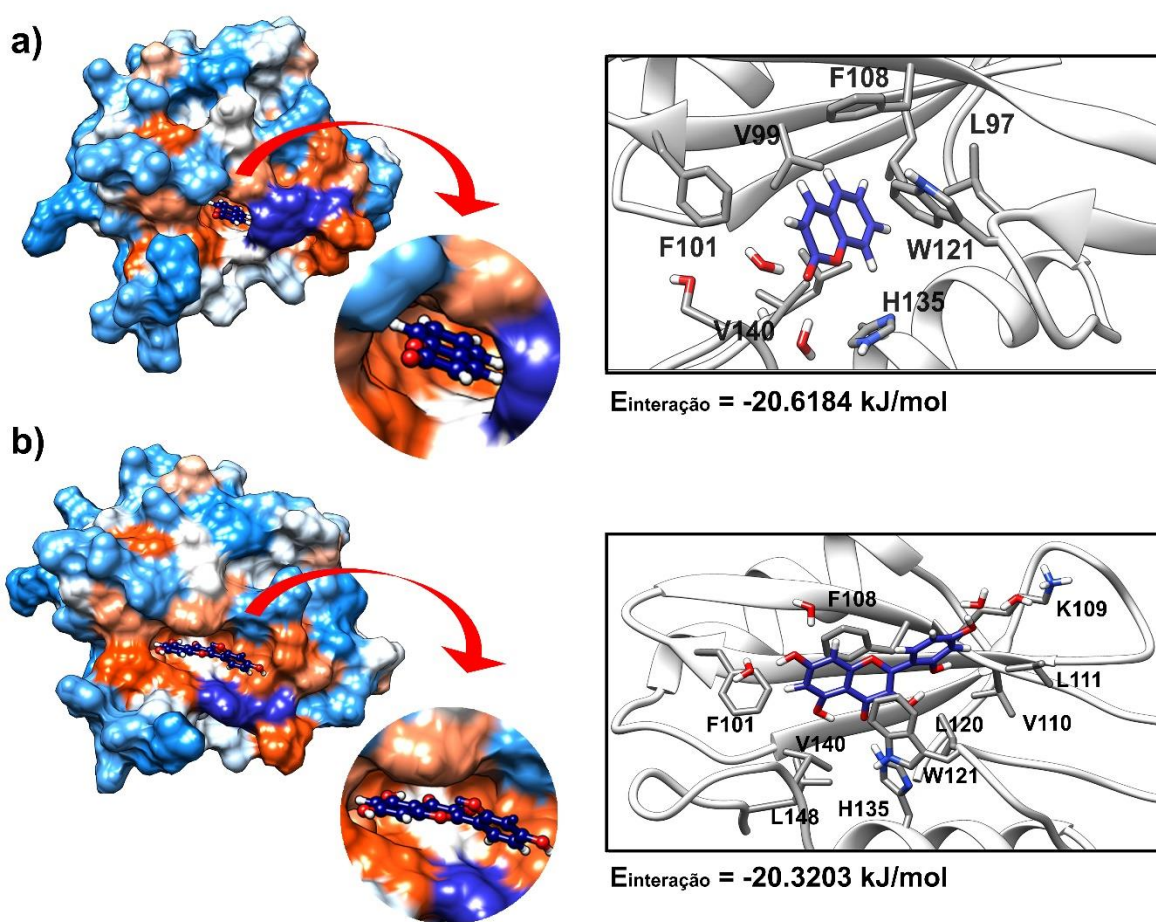
A movimentação dos ligantes para o novo bolsão no domínio SH2 de Grb2 está de acordo com todos os resultados experimentais obtidos até então, demonstrando os possíveis efeitos hidrofóbicos, uma vez que, para o encaixe do ligante na região, moléculas de águas foram repelidas e aconteceram pequenas mudanças conformacionais, contribuindo para o aumento do termo entrópico do sistema.

O resíduo de Triptofano (W121) também estava presente no novo bolsão, e que, devido a interação com o ligante, apresentou alterações conformacionais, corroborando com os resultados de fluorescência, sendo também o responsável pelo ancoramento do ligante ao bolsão, assim como o esperado (Rahuel J, et al., 1996). Por fim, uma análise estrutural de como o ligante se encaixou no sítio, demonstrou que todos os resultados concordam com os epítomos de interação, calculados por RMN-STD.

Uma seleção baseada em um campo de força distinto ao do *docking* molecular buscando convergência é então obtido por meio da ferramenta *g_energy* encontrada no pacote do GROMACS (Pronk et al., 2013), sendo o critério de escolha a conformação com a menor energia de ligação de curto alcance durante a dinâmica molecular, considerada portanto a provável conformação do complexo Grb2-ligantes, como mostrado na Figura 36.

Entretanto, devido toda a parametrização do ligante, aos campos de forças utilizados e o uso do solvente explícito, essa energia retirada da dinâmica molecular não deve ser comparada com os valores obtidos do *docking molecular*, fazendo-se necessário então uma última etapa de *re-docking*, onde todos os processos do *docking* são novamente realizados, levando em consideração agora a nova cavidade formada durante a dinâmica molecular, e permitindo a comparação direta da energia de interação calculada pelo método MM/GBSA com as energias encontradas no *docking* inicial. Para esse processo, ligante e receptor são separados e preparados como descrito nos métodos para *docking* molecular.

Figura 36 – Conformações de menor energia retirada da dinâmica molecular. (A) superfície molecular do domínio SH2 de Grb2. A molécula Cumarina se encaixa no sitio exposto durante a dinâmica molecular, deixando seus resíduos de oxigênios apontados para fora da cavidade e seus carbonos para dentro. Os resíduos próximos a molécula, com distâncias menores que 4Å, são mostrados. (B) A molécula Morina interage no mesmo sitio hidrofóbico, seu tamanho e o jeito de se encaixar faz com essa tenha um maior número de resíduos com distâncias menores que 5Å. Ambas as superfícies são coloridas de acordo com a hidrofobicidade do sítio, indo do laranja para resíduos hidrofóbicos até o azul, para resíduos hidrofílicos.



Fonte: Elaborada pelo autor

2.5.4 Avaliação da interação melhor ranqueada.

Todas as melhores conformações para cada bolsão analisado são mostradas nas tabelas 1 e 2 para os ligantes Cumarina e Morina respectivamente, indicando todos os parâmetros utilizados para o ranqueamento de cada uma dessas conformações.

Tabela 1 – Melhores conformações do ligante Cumarina em Grb2.

Domínio	Pose	E _{Interação} (kJ/mol)	Tem triptofano	Região Hidrofóbica	Epítomos de Interação
Interface dimérica	I	-17.0962	X	X	X
SH2	I	-13.1030	✓	✓	X
SH2 *depois da dinâmica	Única	-20.6184	✓	✓	✓
SH3 N-terminal	I	-11.1468	✓	X	X

Fonte: Elaborada pelo autor

Tabela 2 – Melhores conformações do ligante Morina em Grb2.

Domínio	Pose	E _{Interação} (kJ/mol)	Tem triptofano	Região Hidrofóbica	Epítomos de Interação
Interface dimérica	I	-22.0045	✓	X	X
SH2	I	-21.6184	✓	✓	X
SH2 *depois da dinâmica	Única	-20.3203	✓	✓	✓
SH3 N-terminal	I	-10.9701	✓	X	X
SH3 N-terminal	II	-10.7253	X	✓	X

Fonte: Elaborada pelo autor

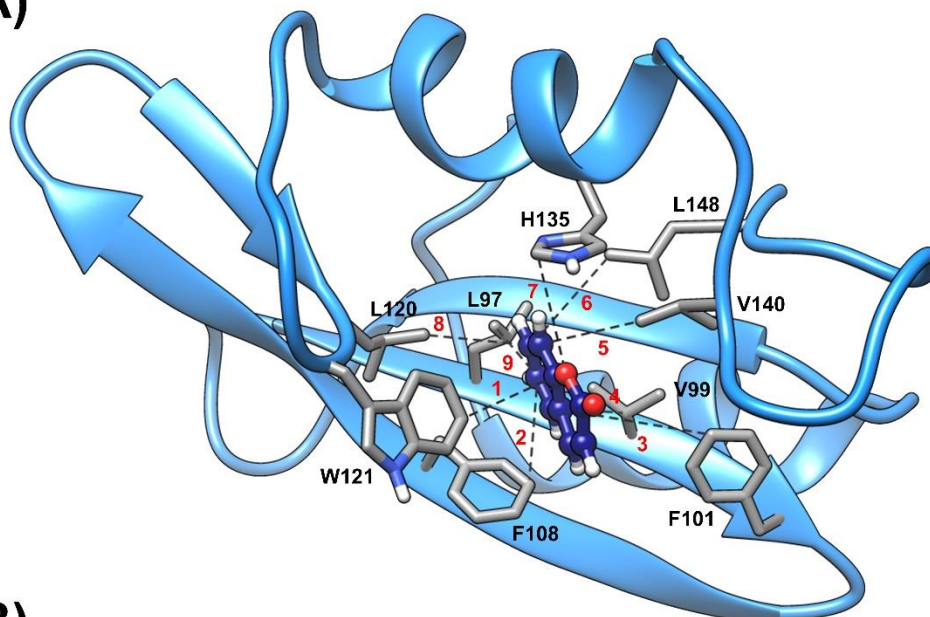
Com todos os resultados de possíveis interações já definidos, o ranqueamento baseado na energia calculada pelo programa de docking molecular e nos parâmetros experimentais de fluorescência e RMN mostram que, a melhor interação acontece no domínio SH2 após a dinâmica molecular, devido as mudanças conformacionais induzidas pelos ligantes e o encaixe no novo bolsão. Nesse bolsão, todos os parâmetros experimentais corroboram com os dados computacionais e a energia de interação, para o caso da Cumarina, também se mostrou como melhor resultado. Para a Morina, apesar dessa energia ter o 3º melhor score, seus valores são bem próximos aos dois melhores, não invalidando a seleção baseada nos parâmetros experimentais.

Para essa região de melhor Score, foi realizado a análise físico-química dos possíveis contatos entre os ligantes e Grb2, apresentado nas Figuras 37 e 38. Para isso, utilizou-se os softwares LigPlot, PLIP e Maestro. Cada resultado individual foi verificado manualmente de forma que o resultado geral é junção de cada um desses resultados.

Para a molécula Cumarina, os efeitos hidrofóbicos foram ocasionados por resíduos apolares, e com distâncias menores que 5Å, sendo esses duas valinas, duas leucinas e uma valina. O resíduo de triptofano também contribuiu para efeitos hidrofóbicos além de seus anéis serem responsáveis pela interação *stacking*. O resíduo de histidina, devido à sua característica de poder estar protonado ou desprotonado por volta de pH 7, pode ocasionar em uma ponte salina entre os oxigênios de Cumarina e seu grupo aromático imidazol. A histidina ainda pode fazer uma interação *stacking* devido a disposição de seu anel em paralelo com os anéis do ligante.

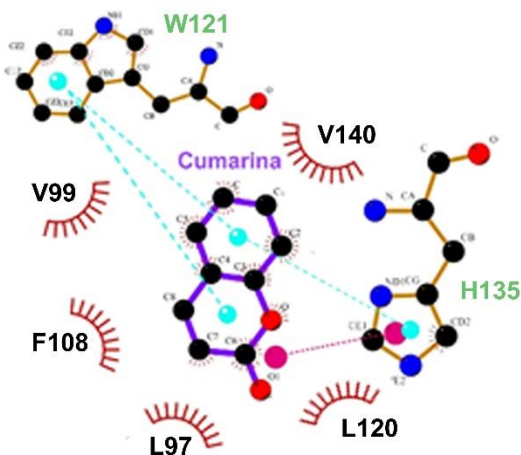
A Morina, apesar da semelhança estrutural com Cumarina, tem diferenças nas interações proteína ligante. Os resíduos responsáveis por efeitos hidrofóbicos são duas leucinas, duas fenilalaninas, uma valina e uma isoleucina, todos esses com distâncias menores que 5 Å. O resíduo de triptofano mais uma vez participa da interação *stacking*, devido à disposição de seus anéis formando ângulos específicos com os anéis da Morina. A histidina pode apresentar uma interação de hidrogênio devido à sua proximidade com o resíduo de oxigênio do ligante, seu anel também pode ser responsável pelo *stacking* com os anéis de Morina. Por fim, uma interação de hidrogênio foi contabilizada entre o átomo de oxigênio da lisina e o oxigênio da Morina.

A)



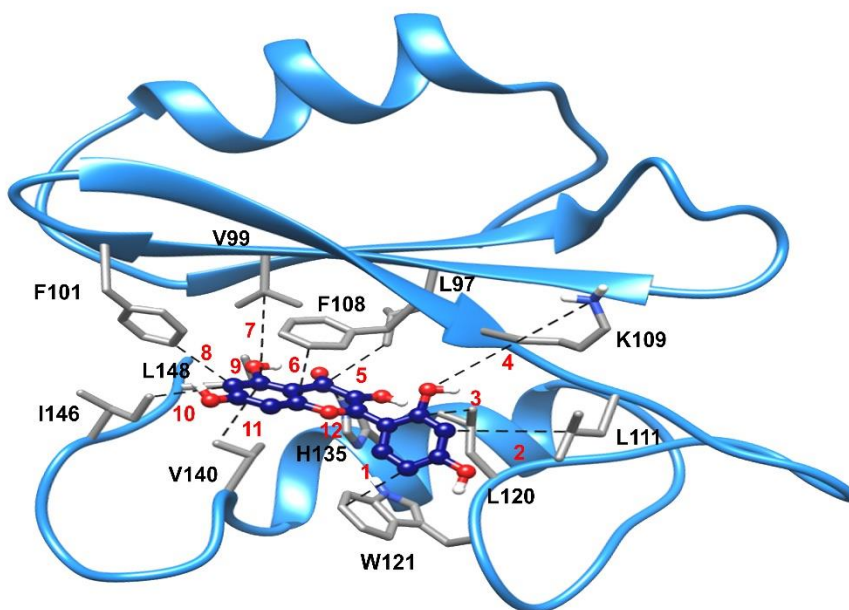
B)

ID	Resíduos	Distância (Å)
1	W121	3.63
2	F108	4.15
3	F101	5.60
4	V99	4.86
5	V140	3.64
6	L148	4.92
7	H135	4.28
8	L97	4.05
9	L120	3.75



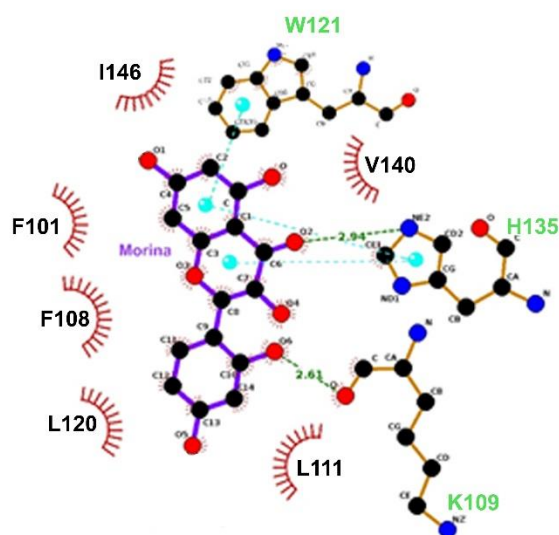
85 | P á g i n a

A)



B)

ID	Resíduos	Distância (Å)
1	W121	3.82
2	L120	4.91
3	L111	3.82
4	K109	8.3
5	L97	4.37
6	F108	3.35
7	V99	5.62
8	F101	4.75
9	L148	6.24
10	L146	3.68
11	V140	3.84
12	H135	3.1



-  **Ligação do Ligante**
-  **Ligação de Hidrogênio e comprimento**
-  **π -Stacking (perpendicular)**

- Resíduos da Proteína envolvidos em efeitos hidrofóbicos
- Resíduos do Ligante envolvidos em efeitos hidrofóbicos
- Ponte Salina

86 | P á g i n a

2.6 Conclusões

Neste trabalho foram avaliados os possíveis sítios de interações da proteína Grb2 com os ligantes Cumarina e Morina, utilizando o *docking* molecular e complementando os resultados com dinâmicas moleculares. Resultados experimentais como RMN e fluorescência no estado estacionário foram utilizados para o refinamento em cada processo realizado e para validação dos resultados obtidos computacionalmente.

Os resultados iniciais de docking molecular mostraram três possíveis bolsões de interação para a Cumarina e quatro para Morina. Para refinamento dos resultados e uma melhor aproximação dos experimentos *in vitro*, foi realizado uma etapa de dinâmica molecular para cada resultado obtido do docking. Essas análises da dinâmica mostraram que, houve a indução de um novo bolsão hidrofóbico no domínio SH2 devido a presença dos ligantes.

O ranqueamento final de todas as estruturas geradas para a interação proteína-ligante gerou um total de 9 conformações possíveis, sendo que, as melhores para ambos os ligantes, foram no novo bolsão formado durante a etapa da dinâmica molecular. Para essas estruturas, foi realizado uma análise físico-química das interações moleculares, evidenciando a predominância de efeitos hidrofóbicos no sítio, assim como possíveis interações envolvendo carga e arranjo dos anéis aromáticos.

Essas indicações de interação no domínio SH2 podem estar relacionadas com mecanismos como, (I) o bloqueio da sua interação com outras moléculas em vias de sinalização, uma vez que SH2 reconhece outros domínios que contenham fosfotirosinas, (II) o equilíbrio de Grb2 entre suas formas dimérica e monomérica (Ahmed et al., 2015). Tais mecanismos correlacionam-se com importantes efeitos fisiológicos, podendo intervir em moléstias como o desenvolvimento tumoral e má formação fetal ocasionados por cascatas de sinalização não reguladas.

A abordagem teórica-experimental se mostrou indispensável para a construção de ideias e novos experimentos de modo complementar, fornecendo uma visão mais ampla do problema abordado, condizente com a área de pesquisa interdisciplinar em que o estudo está inserido.

3. Referências

- Abbas, A. K., Lichtman, A. H., & Pillai, S. (2010). Cellular and molecular immunology. 8ed. Philadelphia: Saunders/Elsevier.
- Ahmed, Z., Lin, C. C., Suen, K. M., Melo, F. A., Levitt, J. A., Suhling, K., & Ladbury, J. E. (2013). Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity. *Journal of Cell Biology*, 200(4), 493–504. <https://doi.org/10.1083/jcb.201204106>
- Ahmed, Z., Timsah, Z., Suen, K. M., Cook, N. P., Lee, G. R., Lin, C.-C., ... Ladbury, J. E. (2015). Grb2 monomer–dimer equilibrium determines normal versus oncogenic function. *Nature Communications*, 6, 7354. <https://doi.org/10.1038/ncomms8354>
- Anfinsen. (1973). ΔG^{TM} CIE : NCES Folding of Protein Chains. *Science*, 181(4096), 223–230.
- Anfinsen, C. B., Haber, E., Sela, M., & White, F. H. (1961). the Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proceedings of the National Academy of Sciences*, 47(9), 1309–1314. <https://doi.org/10.1073/pnas.47.9.1309>
- Angulo, J., & Nieto, P. M. (2011). STD-NMR: Application to transient interactions between biomolecules-a quantitative approach. *European Biophysics Journal*, 40(12), 1357–1369. <https://doi.org/10.1007/s00249-011-0749-5>
- Barkho, S., Pierce, L. C. T., Li, S., Adams, J. A., & Jennings, P. A. (2015). Theoretical insights reveal novel motions in Csk's SH3 domain that control kinase activation. *PLoS ONE*, 10(6), 1–14. <https://doi.org/10.1371/journal.pone.0127724>
- Belov, A. A., Mohammadi, M., & Manuscript, A. (2012). Grb2, a double-edged sword of receptor tyrosine kinase signaling. *Science Signaling*, 5(249), pe49. <https://doi.org/10.1126/scisignal.2003576.Grb2>
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., ... Tasumi, M. (1977). The Protein Data Bank. *European Journal of Biochemistry*, 80(2), 319–324. [https://doi.org/10.1016/S0022-2836\(77\)80200-3](https://doi.org/10.1016/S0022-2836(77)80200-3)
- Biocalorimetry - Applications of calorimetry in the biological sciences. (1996). London: Wiley. (1996). *Biocalorimetry - Applications of calorimetry in the biological sciences*. London: Wiley.
- Boggon, T. J., Li, Y., Manley, P. W., Eck, M. J., Dc, W., Boggon, T. J., ... Eck, M. J. (2012). Crystal structure of the Jak3 kinase domain in complex with a staurosporine analog Crystal structure of the Jak3 kinase domain in complex with a staurosporine analog, 106(3), 996–1002. <https://doi.org/10.1182/blood-2005-02-0707>

- Bolton, E. E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). *Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. Annual Reports in Computational Chemistry* (Vol. 4). Elsevier B.V. [https://doi.org/10.1016/S1574-1400\(08\)00012-1](https://doi.org/10.1016/S1574-1400(08)00012-1)
- Branden, C., Tooze, J. (1999). *Introduction to protein structure*, 2nd ed., Garland publishing, New York.
- Bryngelson, J. D. D., & Wolynes, P. G. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences*, 84(November), 7524–7528. <https://doi.org/10.1073/pnas.84.21.7524>
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3), 167–195. <https://doi.org/10.1002/prot.340210302>
- Calabrò, M. L., Tommasini, S., Donato, P., Stancanelli, R., Raneri, D., Catania, S., ... Ficarra, R. (2005). The rutin/ β -cyclodextrin interactions in fully aqueous solution: Spectroscopic studies and biological assays. *Journal of Pharmaceutical and Biomedical Analysis*, 36(5), 1019–1027. <https://doi.org/10.1016/j.jpba.2004.09.018>
- Caracelli, I., Maganhi, S. H., Stefani, H. A., Gueogjian, K., Tiekink, E. R. T., & Healy, P. C. (2015). Crystal structure of 3-[2-(4-methylphen-yl)ethynyl]-2H-chromen-2-one. *Acta Crystallographica Section E: Structure Reports Online*, 71(2), o90–o91. <https://doi.org/10.1107/S2056989014027790>
- Cavalcanti, A. R. O., & Landweber, L. F. (2003). Genetic Code: What Nature Missed. *Current Biology*, 13(22), R884–R885. <https://doi.org/10.1016/J.CUB.2003.10.052>
- Chan, H. S., & Dill, K. A. (1989). Compact Polymers. *Macromolecules*, 22(12), 4559–4573. <https://doi.org/10.1021/ma00202a031>
- Cheng, A. M., Saxton, T. M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., ... Pawson, T. (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell*, 95(6), 793–803. [https://doi.org/10.1016/S0092-8674\(00\)81702-X](https://doi.org/10.1016/S0092-8674(00)81702-X)
- Chong, Y. P., Chan, A. S., Chan, K. C., Williamson, N. A., Lerner, E. C., Smithgall, T. E., ... Cheng, H. C. (2006). C-terminal Src kinase-homologous kinase (CHK), a unique inhibitor inactivating multiple active conformations of Src family tyrosine kinases. *Journal of Biological Chemistry*, 281(44), 32988–32999. <https://doi.org/10.1074/jbc.M602951200>
- Clementi, C., Jennings, P. A., & Onuchic, J. N. (2001). Prediction of folding mechanism for circular-permuted proteins. *Journal of Molecular Biology*, 311(4), 879–890. <https://doi.org/10.1006/jmbi.2001.4871>
- Clementi, C., Nymeyer, H., & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *Journal of Molecular Biology*, 298(5), 937–953. <https://doi.org/10.1006/jmbi.2000.3693>

- Davidson, D., Bakinowski, M., Thomas, M. L., Horejsi, V., & Veillette, A. (2003). Phosphorylation-dependent regulation of T-cell activation by PAG/Cbp, a lipid raft-associated transmembrane adaptor. *Molecular and Cellular Biology*, 23(6), 2017–28. <https://doi.org/10.1128/MCB.23.6.2017>
- Dhillon, A. S., Hagan, S., Rath, O., & Kolch, W. (2007). MAP kinase signalling pathways in cancer. *Oncogene*, 26(22), 3279–3290. <https://doi.org/10.1038/sj.onc.1210421>
- Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural & Molecular Biology*, 4(1), 10–19. <https://doi.org/10.1038/nsb0197-10>
- Ding, F., Zhao, G., Huang, J., Sun, Y., & Zhang, L. (2009). Fluorescence spectroscopic investigation of the interaction between chloramphenicol and lysozyme. *European Journal of Medicinal Chemistry*, 44(10), 4083–4089. <https://doi.org/10.1016/j.ejmech.2009.04.047>
- Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* 2016, 17, 144.
- Dunn, M. F. (2010). Protein–Ligand Interactions: General Description. In eLS, (Ed.). doi:10.1002/9780470015902.a0001340.pub2
- Ewing, T. J. A., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, 15(5), 411–428. <https://doi.org/10.1023/A:1011115820450>
- Fielding, L. (2007). NMR methods for the determination of protein–ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 51(4), 219–242. <https://doi.org/10.1016/j.pnmrs.2007.04.001>
- Gellman, S. H. Introduction: Molecular Recognition *Chemical Reviews* 1997 97 (5), 1231-1232 <https://doi.org/10.1021/cr970328j>
- Giubellino, A., Burke, T. R., & Bottaro, D. P. (2008). Grb2 signaling in cell motility and cancer. *Expert Opinion on Therapeutic Targets*, 12(8), 1021–33. <https://doi.org/10.1517/14728222.12.8.1021>
- Graves, A. P., Shivakumar, D. M., Boyce, S. E., Jacobson, M. P., Case, D. A., & Shoichet, B. K. (2008). Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. *Journal of Molecular Biology*, 377(3), 914–934. <https://doi.org/10.1016/j.jmb.2008.01.049>
- Howell, B. W., & Cooper, J. A. (1994). Csk Suppression of Src Involves Movement of Csk to Sites of Src Activity, 14(8), 5402–5411.

- Huang, K., Wang, Y. H., Brown, A., & Sun, G. (2009). Identification of N-Terminal Lobe Motifs that Determine the Kinase Activity of the Catalytic Domains and Regulatory Strategies of Src and Csk Protein Tyrosine Kinases. *Journal of Molecular Biology*, 386(4), 1066–1077. <https://doi.org/10.1016/j.jmb.2009.01.012>
- Hubbard, S. R., & Till, J. H. (2000). and F Unction. *New York*, 69(1), 373–98. <https://doi.org/10.1146/annurev.biochem.69.1.373>
- Hunter, T., & Cooper, J. A. (1985). Protein-tyrosine kinases 1, 897–930.
- Jamros, M. A., Oliveira, L. C., Whitford, P. C., Onuchic, J. N., Adams, J. A., Blumenthal, D. K., & Jennings, P. A. (2010). Proteins at work: A combined small angle x-ray scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *Journal of Biological Chemistry*, 285(46), 36121–36128. <https://doi.org/10.1074/jbc.M110.116947>
- Jamros, M. A., Oliveira, L. C., Whitford, P. C., Onuchic, J. N., Adams, J. A., & Jennings, P. A. (2012). Substrate-Specific Reorganization of the Conformational Ensemble of CSK Implicates Novel Modes of Kinase Function. *PLoS Computational Biology*, 8(9), 1–8. <https://doi.org/10.1371/journal.pcbi.1002695>
- Janert, P. (2008). Gnuplot in Action. *Online*, 372. <https://doi.org/10.1002/jhm.840>
- Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., ... Zarrinkar, P. P. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 26(1), 127–132. <https://doi.org/10.1038/nbt1358>
- Karplus, M., Sali, a., & Shakhnovich, E. (1995). Kinetics of protein folding. *Nature*, 373(6516), 665–665. <https://doi.org/10.1038/373665a0>
- Kessel, A. & Ben-Tal, N. (2011). Protein Structure. Introduction to proteins: Structure, Function and motion. Boca Raton (FL); p. 46-132
- Kinoshita, T., Matsubara, M., Ishiguro, H., Okita, K., & Tada, T. (2006). Structure of human Fyn kinase domain complexed with staurosporine. *Biochemical and Biophysical Research Communications*, 346(3), 840–844. <https://doi.org/10.1016/j.bbrc.2006.05.212>
- Koegl, M., Robert, M., Bergman, M., Alitalot, K., & Courtneidge, S. A. (1994). Rapid and efficient purification of Src homology 2 domain-containing proteins : Fyn , Csk and phosphatidylinositol 3-kinase p85, 744, 737–744.
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8), 1011–1021. <https://doi.org/10.1002/jcc.540130812>
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161(2), 269–288. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X)

- Lacy, A. (2004). Studies on Coumarins and Coumarin-Related Compounds to Determine their Therapeutic Role in the Treatment of Cancer. *Current Pharmaceutical Design*, 10(30), 3797–3811. <https://doi.org/10.2174/1381612043382693>
- Ladbury, J. E., & Chowdhry, B. Z. (1996). Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chemistry & Biology*, 3(10), 791–801. [https://doi.org/10.1016/S1074-5521\(96\)90063-0](https://doi.org/10.1016/S1074-5521(96)90063-0)
- Lamers, M. B., Antson, A. A., Hubbard, R. E., Scott, R. K., & Williams, D. H. (1999). Structure of the protein tyrosine kinase domain of C-terminal Src kinase (CSK) in complex with staurosporine. *Journal of Molecular Biology*, 285(2), 713–725. <https://doi.org/10.1006/jmbi.1998.2369>
- Lang, P. T., Brozell, S. R., Mukherjee, S., Pettersen, E. F., Meng, E. C., Thomas, V., ... Kuntz, I. D. (2009). DOCK 6: Combining techniques to model RNA-small molecule complexes. *Rna*, 15(6), 1219–1230. <https://doi.org/10.1261/rna.1563609>
- Lemak, A. S., & Balabaev, N. K. (1994). On The Berendsen Thermostat. *Molecular Simulation*, 13(3), 177–187. <https://doi.org/10.1080/08927029408021981>
- Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, 89(18), 8721–8725. <https://doi.org/10.1073/pnas.89.18.8721>
- Lin, C. C., Melo, F. A., Ghosh, R., Suen, K. M., Stagg, L. J., Kirkpatrick, J., ... Ladbury, J. E. (2012). Inhibition of basal FGF receptor signaling by dimeric Grb2. *Cell*, 149(7), 1514–1524. <https://doi.org/10.1016/j.cell.2012.04.033>
- Lin, X., Ayrapetov, M. K., Lee, S., Parang, K., & Sun, G. (2005). Probing the communication between the regulatory and catalytic domains of a protein tyrosine kinase, Csk. *Biochemistry*, 44(5), 1561–1567. <https://doi.org/10.1021/bi048142j>
- Lindberg, M., Tangrot, J., Oliveberg, M., Tångrot, J., Oliveberg, M., Tangrot, J., ... Oliveberg, M. (2002). Complete change of the protein folding transition state upon circular permutation. *Nature Structural Biology*, 9(11), 818–822. <https://doi.org/10.1038/nsb847>
- Lopez-Gonzalez, J. S., Prado-Garcia, H., Aguilar-Cazares, D., Molina-Guarneros, J. A., Morales-Fuentes, J., & Mandoki, J. J. (2004). Apoptosis and cell cycle disturbances induced by coumarin and 7-hydroxycoumarin on human lung carcinoma cell lines. *Lung Cancer*, 43(3), 275–283. <https://doi.org/10.1016/j.lungcan.2003.09.005>
- Mackerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... Karplus, M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins [†]. *The Journal of Physical Chemistry B*, 102(18), 3586–3616. <https://doi.org/10.1021/jp973084f>
- Maignan, S., Guilloteau, J., Fromage, N., Arnoux, B., Becquart, J., & Ducruix, A. (1995). Crystal structure of the mammalian Grb2 adaptor. *Science*, 268(5208), 291–293. <https://doi.org/10.1126/science.7716522>

- Marshall, M. E., Mohler, L., Edmonds, K., Williams, B., Buffer, K., Ryles, M., ... Cloud, G. (1994). An updated review of the clinical development of coumarin (1 , 2-benzopyrone) and 7-hydroxycoumarin. *Journal of Cancer Research and Clinical Oncology*, 120, 39–42.
- Martins, I., & Jocelyne, A. (2007). Artigo, 30(7), 1597–1605.
- McNemar, C., Snow, M. E., Windsor, W. T., Prongay, A., Mui, P., Zhang, R., ... Weber, P. C. (1997). Thermodynamic and structural analysis of phosphotyrosine polypeptide binding to Grb2-SH2. *Biochemistry*, 36(33), 10006–10014. <https://doi.org/10.1021/bi9704360>
- Meggio, F., Deana, A. D., Ruzzene, M., Brunati, A. M., Cesaro, L., Guerra, B., ... Pinna, L. A. (1995). Different Susceptibility of Protein Kinases to Staurosporine Inhibition: Kinetic Studies and Molecular Bases for the Resistance of Protein Kinase CK2. *European Journal of Biochemistry*, 234(1), 317–322. https://doi.org/10.1111/j.1432-1033.1995.317_c.x
- Menezes, F., Fedorov, A., Baleizão, C., Valeur, B., & Berberan-Santos, M. N. (2013). Methods for the analysis of complex fluorescence decays: sum of Becquerel functions versus sum of exponentials. *Methods and Applications in Fluorescence*, 1(1), 15002. <https://doi.org/10.1088/2050-6120/1/1/015002>
- Meyer, B., & Peters, T. (2003). NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed.*, 42(8), 864–890.
- Mills, J. E., Whitford, P. C., Shaffer, J., Onuchic, J. N., Adams, J. A., & Jennings, P. A. (2007). A Novel Disulfide Bond in the SH2 Domain of the C-terminal Src Kinase Controls Catalytic Activity. *Journal of Molecular Biology*, 365(5), 1460–1468. <https://doi.org/10.1016/j.jmb.2006.10.076>
- Montagner, C. (2007). UNIVERSIDADE FEDERAL DE SANTA CATARINA Programa de Pós-Graduação em Biotecnologia Cristina Montagner ATIVIDADES ANTIFÚNGICA , CITOTÓXICA (CÉLULAS TUMORAIS HUMANAS) E HEMOLÍTICA DE CUMARINAS NATURAIS E SEMI-SINTÉTICAS TUMORAIS HUMANAS) E HEMOLÍTICA DE C.
- Musa, M. A., Cooperwood, J. S., & Khan, M. O. F. (2008). A review of coumarin derivatives in pharmacotherapy of breast cancer. *Current medicinal chemistry* (Vol. 15). <https://doi.org/10.2174/092986708786242877>
- Nada, S., Yagi, T., Takeda, H., Tokunaga, T., Nakagawa, H., Ikawa, Y., ... Aizawa, S. (1993). Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell*, 73(6), 1125–1135. [https://doi.org/10.1016/0092-8674\(93\)90642-4](https://doi.org/10.1016/0092-8674(93)90642-4)
- Naganathan, A. N., & Munoz, V. (2010). Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proceedings of the National Academy of Sciences*, 107(19), 8611–8616. <https://doi.org/10.1073/pnas.1000988107>
- Nelson, D. L., & Cox, M. M. Principios de bioquímica de Lehninger. 5. ed. Porto Alegre: Artmed, 2011.

- Nelson, E. D., & Onuchic, J. N. (1998). Proposed mechanism for stability of proteins to evolutionary mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 95(September), 10682–10686. <https://doi.org/10.1073/pnas.95.18.10682>
- Noel, J. K., Levi, M., Raghunathan, M., Lammert, H., Hayes, R. L., Onuchic, J. N., & Whitford, P. C. (2016). SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Computational Biology*, 12(3), 1–14. <https://doi.org/10.1371/journal.pcbi.1004794>
- Noel, J. K., Whitford, P. C., & Onuchic, J. N. (2012). The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *Journal of Physical Chemistry B*, 116(29), 8692–8702. <https://doi.org/10.1021/jp300852d>
- Okada, M. (2012). Regulation of the Src family kinases by Csk. *International Journal of Biological Sciences*, 8(10), 1385–1397. <https://doi.org/10.7150/ijbs.5141>
- Okada, M., Nada, S., Yamanashi, Y., Yamamoto, T., & Nakagawa, H. (1991). CSK: a protein-tyrosine kinase involved in regulation of src family kinases. *The Journal of Biological Chemistry*, 266(36), 24249–24252. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1722201>
- Oliveira, L. C., Schug, A., & Onuchic, J. N. (2008). Geometrical features of the protein folding mechanism are a robust property of the energy landscape: A detailed investigation of several reduced models. *Journal of Physical Chemistry B*, 112(19), 6131–6136. <https://doi.org/10.1021/jp0769835>
- Omura, S., Iwai, Y., Hirano, A., Nakagawa, A., Awaya, J., Tsuchya, H., ... Masuma, R. (1977). A new alkaloid AM-2282 OF Streptomyces origin. Taxonomy, fermentation, isolation and preliminary characterization. *The Journal of Antibiotics*, 30(4), 275–282. <https://doi.org/10.7164/antibiotics.30.275>
- Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., & Wolynes, P. G. (1996). Protein folding funnels: the nature of the transition state ensemble. *Folding & Design*, 1(6), 441–50. [https://doi.org/10.1016/S1359-0278\(96\)00060-0](https://doi.org/10.1016/S1359-0278(96)00060-0)
- Pereira, W. L. (2015). Ação antiproliferativa do flavanoide morina e do extrato da folha de oliveira (*Olea europaea* L.) contra a linhagem de célula H460. *Revista brasileira de plantas medicinais*, V. 17, n.4, supl.1, p. 798-806, 2015.
- Pierri, C. L., Parisi, G., & Porcelli, V. (2010). Computational approaches for protein function prediction: A combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(9), 1695–1712. <https://doi.org/10.1016/j.bbapap.2010.04.008>
- Pronk, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>
- Rang, H.P., Dale, M.M., Ritter, J.M., Flower, R.J., Henderson, G. *Farmacologia*. 7ª ed. Rio de Janeiro: Elsevier, 2012

- Rognan, D. (2011). Docking Methods for Virtual Screening : Principles and Recent Advances. *Virtual Screening: Principles, Challenges, and Practical Guidelines*, 153–176. <https://doi.org/10.1002/9783527633326.ch6>
- Ronconni, C. (2011). Químicos como arquitetos do mundo molecular 34. | *Ciencia Hoje*, 48, pg. 34-39
- Sakai, H., Chong, K. T., Takeuchi, S., Nakagawa, A., Nada, S., Okada, M., & Tsukihara, T. (2002). Structure of the Carboxyl- terminal Src Kinase , Csk *, 14351–14354. <https://doi.org/10.1074/jbc.C200086200>
- Schiering, N., Knapp, S., Marconi, M., Flocco, M. M., Cui, J., Perego, R., ... Cristiani, C. (2003). Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-Met and its complex with the microbial alkaloid K-252a. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), 12654–12659. <https://doi.org/10.1073/pnas.1734128100>
- Shoichet, B. K., Bodian, D. L., & Kuntzt, I. D. (1992). Molecular Docking Using Shape Descriptors, 13(3), 380–397.
- Sondhi, D., & Cole, P. a. (1999). Domain interactions in protein tyrosine kinase Csk. *Biochemistry*, 38(34), 11147–55. <https://doi.org/10.1021/bi990827+>
- Sun, G., & Budde, R. J. (1999). Mutations in the N-terminal regulatory region reduce the catalytic activity of Csk, but do not affect its recognition of Src. *Archives of Biochemistry and Biophysics*, 367(2), 167–72. <https://doi.org/10.1006/abbi.1999.1253>
- Taketomi, H., Ueda, Y., & Gō, N. (1975). STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION: I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *International Journal of Peptide and Protein Research*, 7(6), 445–459. <https://doi.org/10.1111/j.1399-3011.1975.tb02465.x>
- Tanaka, H., Akagi, K. I., Oneyama, C., Tanaka, M., Sasaki, Y., Kanou, T., ... Ikegami, T. (2013). Identification of a new interaction mode between the Src homology 2 domain of C-terminal Src kinase (Csk) and Csk-binding protein/phosphoprotein associated with glycosphingolipid microdomains. *Journal of Biological Chemistry*, 288(21), 15240–15254. <https://doi.org/10.1074/jbc.M112.439075>
- Thakur, A., Singla, R., & Jaitak, V. (2015). Coumarins as anticancer agents: A review on synthetic strategies, mechanism of action and SAR studies. *European Journal of Medicinal Chemistry*, 101, 476–495. <https://doi.org/10.1016/j.ejmech.2015.07.010>
- Thomas, S. M., & Brugge, J. S. (1997). Cellular Functions Regulated By Src Family Kinases. *Annu. Rev. Cell Dev. Biol*, 13, 513–609. <https://doi.org/10.1146/annurev.cellbio.13.1.513>
- Venugopala, K. N., Rashmi, V., & Odhav, B. (2013). Review on natural coumarin lead compounds for their pharmacological activity. *BioMed Research International*, 2013(Table 1). <https://doi.org/10.1155/2013/963248>

- Voet, Donald; Voet, Judith; Pratt, Charlotte W. Fundamentos de bioquímica: a vida em nível molecular. 4. ed. Porto Alegre: Artmed, 2014
- Wang, J., Oliveira, R. J., Chu, X., Whitford, P. C., Chahine, J., Han, W., ... Leite, V. B. P. (2012). Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proceedings of the National Academy of Sciences*, 109(39), 15763–15768. <https://doi.org/10.1073/pnas.1212842109>
- Whitford, P. C., Miyashita, O., Levy, Y., & Onuchic, J. N. (2007). Conformational Transitions of Adenylate Kinase: Switching by Cracking. *Journal of Molecular Biology*, 366(5), 1661–1671. <https://doi.org/10.1016/j.jmb.2006.11.085>
- Yuriev, E., Agostino, M., & Ramsland, P. A. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, 24(2), 149–164. <https://doi.org/10.1002/jmr.1077>
- Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences*, 89(1), 20–22. <https://doi.org/10.1073/pnas.89.1.20>

Apêndice A

WHAM – *Weighted Histograms Analysis Method*

O método dos múltiplos histogramas ou WHAM (do inglês, *The Weighted Histogram Analysis Method*) é uma generalização do método do histograma simples utilizado para o cálculo de grandezas termodinâmicas. No método do histograma simples é utilizado apenas as simulações em uma única temperatura e montado um histograma de energia que mede a probabilidade de cada uma das energias ocorrer em uma certa temperatura. Para sistemas mais complexos, onde tem-se um maior número de resíduos, uma maior complexibilidade nos graus de liberdade e conseqüentemente um espaço conformacional maior, a simulação em uma única temperatura não é suficiente para acessar todas as conformações possíveis, precisando então do uso de vários histogramas de simulações em diferentes temperaturas.

A função probabilidade para as temperaturas T e T' são dadas pelas seguintes equações:

$$P_{\beta}(E, Q) = \frac{\Omega(E, Q)e^{-\beta E}}{\sum_E \Omega(E, Q)e^{-\beta E}} \quad (\text{eq. A.1})$$

$$P_{\beta'}(E, Q) = \frac{\Omega(E, Q)e^{-\beta' E}}{\sum_E \Omega(E, Q)e^{-\beta' E}} \quad (\text{eq. A.2})$$

Onde β e β' são $1/K_B T$ e $1/K_B T'$ respectivamente e E, Q a densidade de estados. Dividindo a equação A.1 por A.2, temos:

$$P_{\beta'}(E, Q) = \frac{P_{\beta}(E, Q)e^{\beta' - \beta}}{\sum_E e^{-E(\beta' - \beta)}} \quad (\text{eq. A.3})$$

Assim temos a estimativa da quantidade $P_{\beta'}(E, Q)$ a uma temperatura T' a partir de uma simulação feita à temperatura T. considerando que para um conjunto de N simulações, temos a geração de vários histogramas $H_n(E, Q)$, a densidade de estados é dada por:

$$\Omega_n(E, Q) = \overline{H_n}(E, Q) e^{-\beta_n(\vartheta_{\beta_n} - E)} \quad (\text{eq A.4})$$

Onde $\vartheta = -\frac{1}{\beta} \ln Z$ é a energia livre do sistema, sendo Z a função de partição.

Para um conjunto de vários histogramas H_n retirados de diversas simulações em diferentes temperaturas, a densidade de estado real $\Omega(E, Q)$ é a média ponderada da densidade obtida em cada simulação $\Omega_n(E, Q)$

$$\Omega(E, Q) = \sum_{n=1}^N \rho_n \Omega_n(E, Q) \quad (\text{eq. A.5})$$

Os pesos atribuídos ρ_n são obtidos pela minimização de erro de cada histograma utilizado. A probabilidade do sistema assumir a configuração (E,Q) na temperatura T é então dada por:

$$P(E, Q) = \frac{\sum_{n=1}^N H_n(E, Q) e^{\beta_n E}}{\sum_{n=1}^N C_n(E, Q) e^{\beta_n(\beta_n \vartheta - E)}} \quad (\text{eq. A.6})$$

Onde C_n representa o número de configurações na enésima simulação para uma dada temperatura T_n . O método é auto consistente, onde valores arbitrários são atribuídos para a energia livre (ϑ) e partir deste encontra-se ρ , então o caminho inverso é realizado e o processo é repetido diversas vezes até a convergência. Com a probabilidade definida, as grandezas termodinâmicas do sistema são calculadas, por exemplo, podemos calcular a energia livre através da expressão:

$$e^{-\vartheta_n} = \sum P(E, Q) \quad (\text{eq. A.7})$$

Com a energia livre obtida, é necessário ainda sua normalização pela temperatura de transição, definida como sendo a temperatura onde se tem igual população dos ensembles, possibilitando assim a comparação direta de diferentes construções.

Apêndice B

Dinâmica Molecular

O conceito de Dinâmica molecular clássica (MD) foi inicialmente proposto por Alder & Wainwright como um método o qual se estudam os movimentos em um sistema de partículas por simulação. Seus elementos essenciais são o conhecimento do potencial de interação entre as partículas e das equações de movimento clássicos que governam a dinâmica dessas partículas. O potencial de interação deve incluir vários termos que descrevem as interações entre átomos e moléculas. Na DM a força F_i que atua sobre cada partícula i é calculada a partir da derivada da função da energia potencial $E(r_i)$, que descreve a interação entre as partículas, em relação a sua posição r_i , sendo $i = 1, 2, 3 \dots N$, contemplando todos os átomos do sistema a ser simulado.

$$F_i = -\frac{\partial E(r_i)}{\partial r_i} \quad (\text{eq. B.1})$$

A aceleração de cada átomo é obtida dividindo a força F_i calculada pela sua massa m_i :

$$a_i = \frac{F_i}{m_i} \quad (\text{eq. B.2})$$

Substituindo a aceleração calculada no algoritmo de Verlet se determina a posição das partículas em incrementos de tempo δt :

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + a_i(t)\delta t^2 \quad (\text{eq. B.3})$$

Apesar de a predição das novas posições no instante $(t + \delta t)$ sejam computadas somente a partir das posições nos instantes t , $t - \delta t$ e das forças F_i sobre cada partícula no instante t , as velocidades de cada átomo do sistema também são necessárias para o cálculo da energia cinética que, quando somada à energia potencial, dá a energia total do sistema. Elas são obtidas a partir do algoritmo de Verlet para a propagação das velocidades:

$$v_i(t) = \frac{(R_i(t+\delta t) - r_i(t-\delta t))}{2\delta t} \quad (\text{eq. B.4})$$

No algoritmo de Leap-Frog a velocidade é incluída explicitamente na determinação das novas posições e possui uma maior precisão numérica sendo as velocidades calculadas em $(t \pm \Delta t/2)$ e as posições em $(t \pm \Delta t)$. Esse algoritmo introduz pequenas modificações nas equações anteriores, de forma a reduzir as diferenças entre grandes termos, quando as velocidades são calculadas

$$r_i(t + \delta t) = r_i(t) + v_i\left(t + \frac{\delta t}{2}\right) \delta t \quad (\text{eq. B.5})$$

$$v_i\left(t + \frac{\delta t}{2}\right) = v_i\left(t + \frac{\delta t}{2}\right) + a_i(t) \delta t \quad (\text{eq. B.6})$$

Durante a simulação de dinâmica molecular, o processo que requer o mais custo computacional é, em geral, a avaliação das forças para obtenção das acelerações. O tempo gasto no cálculo dessas forças depende da complexidade da função de energia potencial de interação entre as partículas e da complexidade do sistema em estudo. Em sistemas moleculares, as funções potenciais mais realistas são compostos de diversos termos. Na DM são tomadas as derivadas espaciais de cada termo e somadas para se obter a força resultante e a aceleração, sobre cada átomo e a cada novo conjunto de coordenadas. As acelerações são então inseridas no algoritmo utilizado para a predição das novas posições e em seguida todo o processo se repete.

TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, 25 / 05 / 2018

Raphael Vinícius R. Dias

Assinatura do autor.