UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Raphael Vinicius Rodrigues Dias

Mudanças conformacionais e modificações pós-traducionais: um estudo estrutural de mecanismos e comportamentos de sistemas proteicos

São José do Rio Preto
2023

Raphael Vinicius Rodrigues Dias

Mudanças conformacionais e modificações pós-traducionais: um estudo estrutural de mecanismos e comportamentos de sistemas proteicos

Tese apresentada como parte dos requisitos para obtenção do título de Doutor 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", Câmpus de São José do Rio Preto.

Financiadora: CAPES (88882.434371/2019-01)

Orientador: Prof. Dr. Leandro Cristante de Oliveira

Coorientador: Prof. Dr. Fernando Alves de Melo

São José do Rio Preto

2023

D541m

Dias, Raphael Vinicius Rodrigues

Mudanças conformacionais e modificações pós-traducionais: um estudo estrutural de mecanismos e comportamentos de sistemas proteicos / Raphael Vinicius Rodrigues Dias. -- São José do Rio Preto, 2023

132 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Leandro Cristante de Oliveira

Coorientador: Fernando Alves de Melo

1. Enovelamento de Proteínas. 2. Growth factor receptor-bound protein 2 (GRB2). 3. Glicoproteínas. 4. Modificações pós traducionais. 5. Rotas de enovelamento. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Raphael Vinicius Rodrigues Dias

Mudanças conformacionais e modificações pós-traducionais: um estudo estrutural de mecanismos e comportamentos de sistemas proteicos

Tese apresentada como parte dos requisitos para obtenção do título de Doutor 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", Câmpus de São José do Rio Preto.

Financiadora: CAPES (88882.434371/2019-01)

Orientador: Prof. Dr. Leandro Cristante de Oliveira

Coorientador: Prof. Dr. Fernando Alves de Melo

Comissão Examinadora:

Prof. Dr. Leandro Cristante de Oliveira (orientador)
Universidade Estadual Paulista (UNESP/IBILCE)

Prof. Dr. Fabio Ceneviva Lacerda de Almeida
Universidade Federal do Rio de Janeiro (UFRJ)

Prof. Dr. Marcelo Andrés Fossey
Universidade Estadual Paulista (UNESP/IBILCE)

Profa. Dra. Adriana Franco Paes Leme
Centro Nacional de Pesquisa em Energia e Materiais (CNPEM)

Dr. Antonio Bento de Oliveira Junior
Rice University - Center for Theoretical Biological Physic

São José do Rio Preto
27 de Janeiro de 2023

Agradecimentos

Agradeço primeiramente ao meu orientador Leandro Cristante, que me incentivou e aconselhou cada passo dentro da universidade. Desde a graduação, quando eu cheguei perdido na sala dele precisando de um projeto para um tcc, projeto esse que desencadeou toda carreira acadêmica até então. Muito obrigado pelos conselhos, conversas, confiança e paciência.

Também agradeço ao meu coorientador Fernando Alves de Melo, que também desde a graduação me acolheu em seu laboratório, sempre com ótimas discussões e MUITO trabalho. Muito obrigado também professor, você foi e é fundamental para essa caminhada trilhada até aqui! Pode ter certeza que nunca vou esquecer de "mexer o doce"!!!

Agradeço à minha família, os quais sempre estiveram junto nessa caminhada, me escutando, me incentivando e sempre provendo ótimas condições. Muito obrigado por toda paciência, compreensão e por sempre estarem presentes. Artur, você está incluso aqui viu !

Aos meus amigos: Larissa e Carolina (DoPreDin), que acompanharam de perto cada etapa desse trabalho, sempre ajudando e discutindo muita ciência, mesmo na pandemia com nossas reuniões semanais e apresentações que só enriqueceram o nosso conhecimento! Ps. Larissa, você foi fundamental nesses ultimos anos, como você mesmo fala, acho que eu teria surtado se não fosse você!

Os menino Renan e Muliro (Murilo), vocês foram extremamente importantes para esse trabalho. Desde o artigo "É TUDU PROS MININU"até as outras coisas que vocês sempre me ajudaram! As nossas discussões sempre muito valiosas, apesar das risadas e palhaçadas. Os menino Renan, que mesmo na pandemia estava sempre aqui trabalhando junto, almoçando e jogando também, com direito a uma etapa stremer scientifico que vivemos. E Muliro, apresentar o zeldinha nos ultimos meses de doutorado foi sacanagem !

Gi e Naza, muito obrigado por tudo! Cada discussão sempre tem uma tonelada de idéias, é até dificil de acompanhar. Como eu sempre digo, enquanto as idéias estão vindo, Giovana e Naza já estão voltando com o experimento pronto e novas coisas! Gi, agradeço muito pela confiança, valeu a pena ter te atormentado uns meses, até você decidir se juntar a nós! E Naza, nunca vou me esquecer do trote "sem ser minha veterana".

João Paulo (ROOOOSA JUÃO), Coragem, Tharita (doutorado da depressão vai ficar pra sempre!), Ingrid e Isa todos sempre muito solicitos, muito obrigado por tudo. Ishabella que desde 2010 me atura, entre outros que passaram e deixaram suas marcas. Meu muito obrigado !

Agradeço também as agências de fomento CAPES e FAPESP, pelo apoio financeiro que proporcionou a realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - (8882.434371/2019-01).

Por fim, agradeço ao apoio do Instituto de Biociências, Letras e Ciências Exatas (IBILCE - Unesp), ao departamento de física pela infraestrutura fornecida e aos técnicos, Barbosa e Bruno. Podem ter certeza, que nada funcionaria sem vocês por ai !

"I know that new situations can be intimidating. You're lookin' around and it's all scary and different, but y'know, meeting them head-on, charging into 'em like a bull — that's how we grow as people.

Wubba Lubba Dub Dub"

(Rick Sanchez)

Resumo

O estudo de mecanismos no enovelamento de proteínas permite extrair importantes informações sobre o funcionamento de sistemas biológicos, assim como prospectar condições anômalas devido modificações como mutações e glicosilações. Neste trabalho, utilizamos modelos simplificados para estudar dois problemas diferentes envolvendo o enovelamento de proteínas: (I) a indicação de estados metaestáveis da proteína GRB2, apontados por meio de avaliações experimentais e (II) a influência das modificações pós-traducionais (glicosilações) no perfil de enovelamento. A GRB2 é uma proteína multidomínios relacionada com interações e regulação de vias dentro da célula, estruturalmente é formada por dois domínios SH3s flanqueados por um domínio SH2. Os estudos computacionais permitem descrever possíveis caminhos para o enovelamento, inclusive com a diferenciação na ordem de formação dos domínios. Complementarmente, mostramos a relação entre deleção do domínio SH3 N-terminal para com o seu relevo e funil de energia, validando e linkando com resultados de estabilidade térmica já estabelecidos previamente. Ao analisar as diferenças entre o enovelamento de cada domínio separadamente e constituindo a proteína completa, foi verificado pequenas diferenças no processo onde os efeitos da comunicação entre domínios fica evidenciada. No estudo da influência das modificações pós-traducionais (glicosilações) no perfil de enovelamento para proteínas depositadas no PDB, modelamos as glicosilações de acordo com nossos parâmetros de ajuste, simulamos o processo de enovelamento e calculamos aspectos termodinâmicos estatísticos como Calor Específico e Perfil de Energia Livre para cada sistema. Dentre seus resultados, notamos que a presença da glicosilação pode influenciar diretamente no enovelamento em alguns casos específicos, alterando graus de flexibilidade, estabilidade térmica e a importância de alguns resíduos para construção do "ensemble" de transição. Descrever esses mecanismos é um ponto de partida para diversos outros estudos e metodologias que interfiram diretamente nos problemas relacionados com essas proteínas, como por exemplo transporte e desenho racional de fármacos, mutações para especificidade de interação e estabilidade.

Palavras chaves: Enovelamento de proteínas. GRB2. Mudanças Conformacionais. Glicosilações.

Abstract

The study of mechanisms of protein folding allows us to extract important information about functions of biological systems, as well as to prospect anomalous conditions due to modifications such as mutations and glycosylations. In this work, we used simplified models to study two different problems involving protein folding: (I) evidences of metastable states of the GRB2 protein, indicated by means of experimental evaluations and (II) the influence on the energetic profile of folding in post-translational modifications (glycosylations). Growth factor receptor bound protein 2 (GRB2) is a multidomain protein related to interactions and regulation of pathways within the cell, structurally it is formed by two SH3s domains flanked by an SH2 domain. Computational studies allow us to describe possible paths for folding, including differentiation in the order of formation of domains. In addition, we show the relationship between deletion of the N-terminal SH3 domain and energy funnel, validating and linking it with previously established thermal stability results. When analyzing the differences between the folding of each domain separately and constituting the complete protein, it was verified small differences in the process where the effects of communication between domains is evidenced. In the study of the influence of post-translational modifications (glycosylations) on the folding profile for proteins deposited in the PDB, we modeled the glycosylations according to our adjustment parameters, simulated the folding process and calculated statistical thermodynamic aspects such as specific heat at constant volume and Energy Profile Free for each system. Among their results, we noticed that the presence of glycosylation can directly influence the folding in some specific cases, changing degrees of flexibility, thermal stability and the importance of some residues for the construction of the transition ensemble. Describing these mechanisms is a starting point for several other studies and methodologies that directly interfere with problems related to these proteins, such as transport and rational drug design, mutations for interaction specificity and stability.

Keywords: Protein Folding. GRB2. Conformational Changes. Glycosylation.

Lista de Figuras

1 Estruturas da proteína. A) Exemplo de estrutura primária da proteína, representada apenas pela sequência de aminoácidos que formam a proteína. B) Hélice alfa colorida em laranja e folhas Beta em ciano representando as estruturas secundárias. C) Estrutura terciária com folhas betas coloridas em ciano, alfa hélices em laranja e *random coils* cinzas formando a estrutura dobrada de um monômero proteico. D) Dois monômeros ligados entre si, formando-se uma estrutura dimérica representando o que seria uma estrutura quaternária. As estruturas utilizadas para a construção da imagem foram retiradas de banco de dados do PDB, utilizando o código 1Z7X. 13

2 Representação do funil de energia. Ilustração do funil de estruturas do processo de enovelamento de uma proteína ao longo da coordenada de reação Q. Na figura são apresentadas as componentes Entropia S(Q) e Energia E(Q). As estruturas desenhadas à direita representam os respectivos ensembles das condições nomeadas abaixo com um ponto indicativo da conformação no funil. 17

3 Diagrama do Estado de Transição (TSE). O diagrama ilustra a energia dos estados desenovelados, de transicão e enovelados. Em casos onde o resíduo mutado não interfere na formação de TSE, os valores de ϕ são próximos de 0 e representados pelo esquema da esquerda. Quando o resíduos mutado interfere na formação de TSE, ϕ é próximo de 1 e o esquema que representa essa situação está a direita do diagrama. 21

4 Proteína GRB2. A) Sequência de aminoácidos da proteína GRB2 monomérica, as sequências são coloridas de acordo com o domínio pertencente: SH3 $N_{TERMINAL}$ em azul, SH2 em roxo e SH3 $C_{TERMINAL}$ em laranja. Em destaque também são mostradas as estruturas secundárias seguindo a construção da estrutura 3D. B, C e D) Domínio SH3 $N_{TERMINAL}$, SH2 e SH3 $C_{TERMINAL}$ respectivamente. Todos eles são coloridos de acordo com as estruturas secundárias, sendo em verde as folhas betas e laranja as hélices alfas. 26

5 Representação do mecanismo de regulação da proteína quinase FGFR2 através da proteína adaptadora GRB2. A proteína FGFR2 é representada mimeticamente pelas barras cinzas simbolizando sua região citoplasmática e com seu domínio quinase 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 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 sinalização.. . . . 28

6 Perfil termodinâmico da proteína GRB2. A) Mapa de contatos calculados utilizando o algoritmo CSU. Cada ponto no gráfico representa o contato do resíduo em X com o resíduo em Y. Os contatos são coloridos de acordo com o domínio onde é encontrado, sendo azul SH3 NTERMINAL, Roxo SH2 e Laranja SH3 CTERMINAL. Em Cinza também são representados os contatos entre os domínios. B) Calor Específico a volume constante e fração de contatos, mostrando a temperatura de transição (dado em unidades reduzidas). C) Energia Livre em função da coordenada de reação dos contatos nativos (Q). A curva mostra diversos mínimos de energia entre os estados enovelados e desenovelados, mostrando um enovelamento que passa por estados intermediários. O valor da barreira de energia também mostra que a última transição (intermediario/Nativo) é a mais custosa energeticamente em relação as outras transições 31

7 Superfície de energia e clusterização de estruturas. A) Superficie de energia livre construída utilizando a correlação das coordenadas de reação dos contatos nativos (Q) e RMSD. Essa superfície é colorida de acordo com a energia livre de cada estado, mostrando mais uma vez 3 estados intermediários entre os estados nativo e desenovelado. Os números destacados (1,2 e 3) foram utilizados para a clusterização das estruturas pertencentes aqueles estados, sendo representados via tubos em 3D, coloridos utilizando a paleta "*rainbow*" do vermelho para o azul. 32

8 Rota de enovelamento e Formação de contatos ao longo do enovelamento. (A) Perfil de Energia Livre. (B) Formação dos contatos separados por domínios. Cada cor representa os contatos dentro de um determinado domínio (vermelho - SH3N, preto - SH2 e verde SH3C), as porcentagens calculadas mostram quanto de cada um desses domínios estão formados de acordo com a coordenada de reação dos contatos nativos. (C) Rota de enovelamento R(Q) que é calculada acompanhando a coordenada de reação Q (contatos nativos) ao longo da dinâmica. Os dois picos mostram que nesses pontos há uma recorrência na formação de uma determinada estrutura. 34

9 Probabilidades de formação de contatos e Funil de estruturas. Funil de estruturas, construído utilizando as estruturas retiradas da dinâmica de enovelamento. Cada ponto representa uma configuração conformacional e as distâncias entre os pontos determinam a similaridade entre elas. Esse funil é colorido de acordo com a fração normalizada dos contatos nativos (QGRB2). Para complementar os resultados, é mostrado os mapas de contato, destacando os ensembles com cada mapa correspondente. Os mapas foram coloridos de acordo com a probabilidade de formação de contato (Pij). passando de poucos contatos formados e domínios desdobrados (azul) para contatos formados e domínios dobrados (vermelho). . . 35

10 Caminhos de mínima energia. A) Superfície de energia livre construída utilizando os contatos de cada domínio de GRB2 como coordenadas de reação. Os valores de Energia Livre seguem a escala da barra de cores apresentada abaixo de 0 (preto) a 8 (amarelo) normalizados pela temperatura do estado de transição (Ts). B) Espaço de fase conformacional de GRB2 visualizado através do ELViM em função dos contatos nativos Q, e com as diferentes rotas destacadas em vermelho e azul. 36

11 Deleções de domínios e efeitos nos caminhos de enovelamento. Estruturas usadas para dinâmica molecular, coloridas de acordo com a estrutura secundária, com o verde representando as folhas$-\beta$ e laranja para as α-hélices. Os funis de energia mostram que a deleção do domínio SH3 $N_{TERMINAL}$ afeta os ensembles de estruturas, armadilhando a dinâmica no mínimo em que há apenas o $C_{TERMINAL}$ formado, desestabilizando a formação do domínio SH2. Já a deleção do SH3 $C_{TERMINAL}$ não afeta as rotas ou desestabiliza a formação de domínios. 38

12 Parâmetros termodinâmicos dos domínios de Grb2. Na parte inferior, está sendo mostrado a curva de calor específico pela temperatura em unidades reduzidas, os valores destacados para cada curva referem-se à temperatura do pico e também ao valor utilizado para cálculo da curva de energia livre mostrado na parte superior da imagem. A coordenação de reação utilizada são os contatos nativos (Q), o qual permite acompanhar as estruturas formadas ao longo do enovelamento. Também são mostrados cada domínio coloridos por suas estruturas secundárias: em roxo as folhas betas, laranja hélice alfas e cinza random coils, e seus nomes com cores referentes a cada curva nos gráficos de CV e Energia Livre. 39

13 Rotas de enovelamento dos domínios de GRB2. Estão sendo mostradas as rotas dos 3 diferentes domínios de GRB2 isolados e fazendo parte do monômero. Para os casos de SH3 $N_{TERMINAL}$ e SH2 o perfil no início das curvas diferem, tendo um valor mais alto para as curvas que representam os domínios ligados (monômero). Isso é explicado peloa fato de que as pontas C-terminais estão presas quando nessa forma, diminuindo o grau de liberdade das possibilidades e consequentemente aumentando a especificidade na formação das estruturas. Já para SH3 $C_{TERMINAL}$, há uma alteração nessa especificidade no intervalo médio dos contatos nativos. 41

14 Heterogeneidade na formação do domínio SH3 $C_{TERMINAL}$. A) Probabilidade na formação dos contatos nativos com 50,60 e 70% da estrutura formada utilizando as dinâmicas do monômero de GRB2 e B) utilizando as dinâmicas do domínio isolado. Nota-se aqui a formação conjunta das folhas betas do domínio SH3 $C_{TEMINAL}$ quando na forma monomérica de GRB2. Quando o domínio encontra-se isolado, essa formação não é heterogênea, sendo necessário a construção de algumas estruturas para depois finalizar a última beta. Os quadrados destacados no calculos de 60% podem ser utilizados para uma melhor visualização do fenômeno. As estruturas 3D nos cantos inferiores de cada gráfico ajudam a visualizar, sendo que cada linha entre os átomos representa as interações com probabilidade maior que 50%. 42

15 Sítios de ligação de N-glicosilações e O-glicosilações. Na figura estão representados os dois tipos mais comuns de ligação dos glicanos em proteínas. . 46

16 Estrutura da Proteína CD2. O domínio de adesão da proteína CD2 humana é representado juntamente com sua glicosilação. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de ASN65 em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho. 49

17 Estrutura da Proteína IgG-FC. Também conhecida como Imunoglobulina, IgG-FC é representado juntamente com sua glicosilação. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de Asn em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho. 52

18 Estrutura da Proteína Erns. A proteína viral Erns é representada juntamente com suas glicosilações. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de Asn em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho. 53

19 Modelagem da glicosilação com o SBM. Um esquema de glicosilação simples é mostrado, com apenas 4 anéis sendo representados. As esferas nomeadas com C representam os carbonos da proteína em que está ligado o açúcar, as nomeadas com G representam o próprio açúcar e os índices indicam os resíduos ou anéis de sacarídeos. 54

20 Adição dos parâmetros na topologia da proteína. Para um esquema como o que foi mostrado na figura 19, as novas combinações de resíduos com a adição do glicano é representada. Essas combinações representam os novos parâmetros adicionados no arquivo de topologia e que representam a presença do glicano na estrutura proteica. 55

21 Modelagem das proteínas glicosiladas. Na parte superior estão representadas as proteínas coloridas por suas estruturas secundárias (hélice alfa em laranja, folhas betas em azul e random coil em cinza) juntamente com as glicosilações em laranja. Na parte inferior, as proteínas já modeladas por seus carbonos alfa e glicosilações, em amarelo estão representados as ASN em que estão ligados as glicosilações em laranja. 56

22 Calor Específico a volume constante. (A) Calor específico x Temperatura em unidades reduzidas das proteínas estudadas. (B) Sobreposição das curvas contínuas e tracejadas mostrada em A. (C) Fração de contatos nativos para cada uma das proteínas estudadas. Para todos os 3 casos, as curvas contínuas e tracejadas correspondem às estruturas sem e com glicosilações respectivamente. Com a adição das glicosilações, as proteínas PDB ids: 1GYA e 4DW5 tiveram suas curvas de Cv e Fração de Contatos desviadas para a direita, evidenciando um possível aumento em sua estabilidade térmica. As sobreposições mostram que não há um alargamento das curvas, descartando hipóteses de cooperatividade provindas da adição de glicosilações. 58

23 Energia Livre para os casos glicosilados e deglicosilados. (A, B) Perfil de energia livre para a coordenada de contatos nativos e RMSD respectivamente. (C) Superfície de energia utilizando a correlação de ambas as coordenadas. Para todos os casos, as curvas contínuas e tracejadas correspondem às estruturas sem e com glicosilações respectivamente. As curvas em preto, representando a proteína CD2, mostra que a adição da glicosilação afeta a barreira de energia livre (vista apenas para a coordenada RMSD) e é uma transição de dois estados. Para a proteína Imunoglobulina (curvas em vermelho), não há alterações nas barreiras de energia e é uma transição passando por um estado semi-enovelado devido a presença dos dois diferentes domínios. Por fim, as curvas em azul representam a proteína Erns em que há um estado intermediário facilitado pela adição de glicosilação e um aumento na barreira para seu desenovelamento. 59

24 Valores ϕ. (A,B e C) Curvas das proteínas com código PDB 1GYA, 1H3W e 4DW5 respectivamente. Nessas, as curvas com os valores phi com e sem glicosilação são sobrepostas, mostrando a influência da glicosilação em alguns contatos. As relações entre aumento e diminuição nos valores phi são mais evidenciadas nos gráficos (A) e (C), destacando-se alguns possíveis resíduos de importância para o estado de transição, como mostrado em (D). 61

25 Perfil de flutuação (RMSF). Na esquerda são apresentados os perfis das proteínas deglicosiladas e na direita as proteínas glicosiladas. As cores são representadas pelas temperaturas das quais foram retirados o perfil de flutuação. Nota-se que a glicosilação teve como efeito preencher os estados com flutuações intermediárias. 62

26 Entropia e Entalpia do sistema. Os valores de entalpia e entropia estão sendo mostrados para as três proteínas de interesse. (A,B e C) separam os resultados das proteínas 1GYA, 1H3W e 4DW5 respectivamente. Cada gráfico contém as curvas de Entalpia e entropia, diferenciando-se pela linha tracejada e contínua. E também para cada gráfico é apresentado seus valores tanto na sua forma deglicosilada quanto glicosilada, representados pelas cores preta e vermelha. 64

Lista de Abreviaturas

DM Dinâmica Molecular

α Alfa

Cα Carbono Alfa

GRB2 *Growth Factor Receptor Bound proteín 2*

P_ij Probabilidade de contato entre i e j

GROMACS *Groningen Machine for Chemical Simulations*

SMOG *Structure-based Models for Biomolecules*

ELViM *Energy Landscape Visualization Method*

SH3 *SRC Homology 3 Domain*

SH2 *SRC Homology 2 Domain*

Ans Asparagina

Ser Serina

THR Treonina

Sumário

1 Introdução **12**

1.1 Estrutura e função das proteínas 12

1.2 Breve histórico e fundamentação da Teoria de enovelamento de proteínas: de Anfinsen a superfície de funil 15

2 Metodologia **18**

2.1 Modelos baseados em estrutura 18

2.2 Coordenadas de Reação 19

2.3 Valores *phi* 20

2.4 Rotas de Enovelamento 22

2.5 Energy Landscape Visualization Method (ELViM). 22

2.6 Simulações Computacionais 23

2.7 Trabalhos adicionais e complementares 24

3 *Growth Factor Receptor Bound 2 (GRB2)* **26**

3.1 Introdução 26

3.2 Objetivos 29

3.3 Resultados e Discussões 30

3.3.1 Perfil de Enovelamento da proteína Grb2 30

3.3.2 Formação de contatos e especificidade do enovelamento 33

3.3.3 Funil de energia e estados semi-enovelados 34

3.3.4 Deleção de domínios e estabilidade termodinâmica 37

3.3.5 Perfil termodinâmico dos domínios de GRB2 38

3.4 Conclusões 42

4 Glicoproteinas **46**

4.1 Introdução às Glicoproteínas 46

4.2 Sistemas alvo de estudo 48
4.3 Modelagem do sistema 53
4.4 Objetivos 56
4.5 Resultados e discussões 57
4.5.1 Perfil de Energia Livre dos sistemas 57
4.5.2 Avaliação das influências nos mecanismos de enovelamento por meio dos valores-ϕ 60
4.6 Flutuação média dos resíduos 61
4.7 Entalpia e Entropia das transições 63
4.8 Conclusões 65

Referências **66**

A Trabalhos publicados no período do doutorado **73**

B Trabalhos Submetidos e em preparação. **112**

1 Introdução

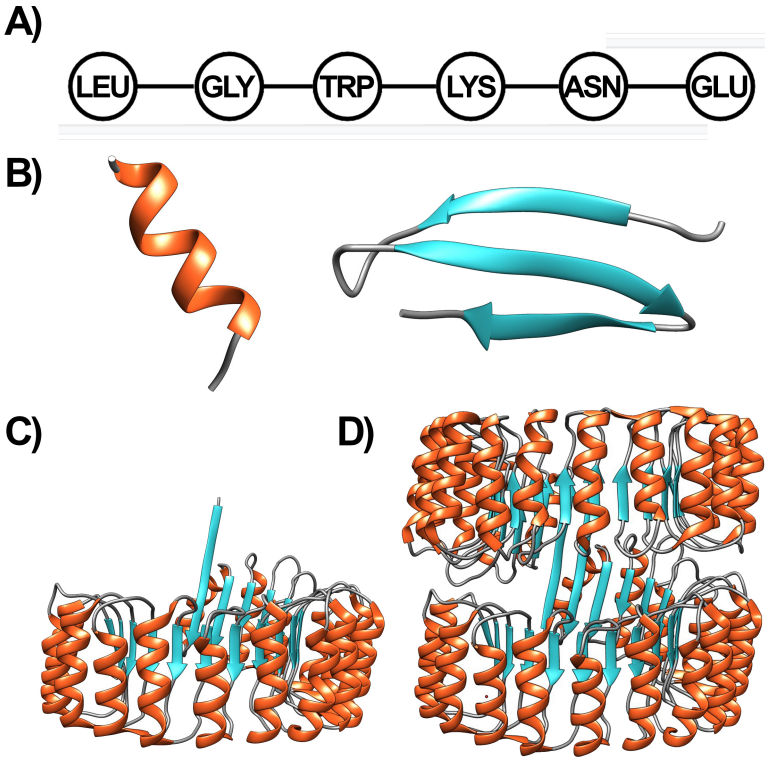
1.1 Estrutura e função das proteínas

Proteínas são macromoléculas responsáveis por grande parte dos processos biológicos em organismos vivos apresentando uma gama de funções tais como: regulação (controlando as condições intra e extracelulares), transporte, recepção de sinais, construção de estruturas celulares (como parte dos músculos e outros tecidos) ou convertendo energia química em energia mecânica (1, 2). O funcionamento proteico e sua importância celular proporciona questionamentos sobre os entendimentos desses processos nos âmbitos físico e químico.

Proteínas são comumente formadas por 20 diferentes tipos de aminoácidos, com exceção a alguns organismos que podem apresentar 22 diferentes tipos de aminoácidos (3), sem considerar ainda os aminoácidos sintéticos (4). Estruturalmente os aminoácidos são constituídos em sua cadeia principal por um carbono, conhecido como carbono alfa (Cα) fazendo quatro ligações covalentes: uma com um átomo de hidrogênio (H), outra com um grupo amina (NH2), outra com um grupo carboxila (COOH) e outra com um grupo radical (R) o qual é responsável por distinguir um aminoácido do outro (1, 2).

A sequência de aminoácidos ligados (resíduos de aminoácidos) covalentemente formando uma cadeia polipeptídica é o que chamamos de estrutura primária de uma proteína. Os ângulos formados pelas ligações peptídicas entre os resíduos conferem padrões que refletem em estruturas locais regulares, como as hélices alfas, folhas betas, entre outras, que por sua vez formam o conjunto conhecido como estruturas secundárias. O rearranjo espacial tridimensional dos aminoácidos já com as estruturas secundárias formadas é chamado de estrutura terciária de uma proteína. Essa estrutura é mantida por interações não covalentes entre aminoácidos distantes na cadeia como por exemplo: ligações de hidrogênios, pontes dissulfeto, pontes salinas e interações hidrofóbicas. Por fim, as estruturas quaternárias são formadas quando a proteína possui duas ou mais subunidades (estruturas terciárias) que se agrupam tridimensionalmente, também conhecidas como dímeros, trímeros, etc. (1, 2). Esses arranjos tridimensionais podem ser resolvidos por técnicas experimentais (por exemplo cristalografia ou Ressonância Magnética Nuclear) e geralmente encontram-se depositados em bancos de dados de proteínas como o Protein Data Bank. (5)

Figura 1: Estruturas da proteína. A) Exemplo de estrutura primária da proteína, representada apenas pela sequência de aminoácidos que formam a proteína. B) Hélice alfa colorida em laranja e folhas Beta em ciano representando as estruturas secundárias. C) Estrutura terciária com folhas betas coloridas em ciano, alfa hélices em laranja e *random coils* cinzas formando a estrutura dobrada de um monômero proteico. D) Dois monômeros ligados entre si, formando-se uma estrutura dimérica representando o que seria uma estrutura quaternária. As estruturas utilizadas para a construção da imagem foram retiradas de banco de dados do PDB, utilizando o código 1Z7X.



Fonte: Elaborada pelo autor (2023).

Inicialmente acreditava-se que a sequência de aminoácidos e consequentemente a estrutura (rearranjo) final de uma proteína era única e determinística para a sua função. Entretanto, novos estudos mostram que este paradigma de estrutura-função não é absoluto, pois nas últimas décadas se tem evidenciado a importância de sua dinâmica, tanto para interação com outras moléculas como até mesmo para fins de suas atividades catalíticas (6, 7).

O avanço das técnicas e melhoramento nas tecnologias nos estudos de macromoléculas, tanto teórico quanto experimental, permite entender que há diversos fatores que estabelecem a atividade de uma proteína, como por exemplo interações com outras macromoléculas e rearranjos conformacionais. Motores biológicos são exemplos de proteínas que dependem

das mudanças conformacionais para que ocorram as suas atividades (8–10). As proteínas intrinsecamente desordenadas (IDPs) que não apresentam uma forma tridimensional fixa mas apresentam funções celulares são exemplos que mostram que a função depende da dinâmica estrutural e não somente da estrutura (11).

Não menos importante, há também as conhecidas modificações pós-traducionais (do inglês, Post-translational modification – MPTs), que são eventos de processamento covalente que mudam as propriedades das proteínas por clivagem proteolítica ou por adição de um grupo químico a um ou mais aminoácidos (12). São conhecidos mais de 200 tipos de mudanças, encontradas em todas as espécies e estão presentes em pelo menos 80% das proteínas eucariotas. Dentre as MPTs possíveis, alguns exemplos são: fosforilação, acetilação, metilação, glicosilação, pontes de dissulfeto, sulfatação e ubiquitinação. E dentre ao grande número de atividades, destaca-se a estabilidade proteica, interações celulares, secreção e funções regulatórias (12, 13). Dentre todas essas MPTs apresentadas, destaca-se aqui as glicosilações em proteínas, que será abordada nos próximos capítulos deste trabalho.

Um mecanismo comum, que envolve todos os processos destacados até então, é o enovelamento da proteína. Esse mecanismo é o responsável por levar a macromolécula do seu estágio inicial (estrutura primária) até seu estado funcional (estruturas terciárias ou quaternárias, dependendo do sistema alvo). Ainda, há indicativos na literatura mostrando que as anormalidades desse processo podem estar diretamente ligadas a doenças degenerativas como Alzheimer, Huntington e Parkinson (14). Assim, a compreensão de alguns dos mecanismos biomoleculares ao longo do processo de enovelamento das proteínas tem sido um desafio para pesquisadores da área, estando além da relação estrutura-função (15). Estudos dessa natureza, por vezes, envolvem cientistas de diversas áreas do conhecimento por seu contexto interdisciplinar, evidenciando que a compreensão dos princípios e especificidades que governam tais mecanismos não são triviais.

1.2 Breve histórico e fundamentação da Teoria de enovelamento de proteínas: de Anfinsen a superfície de funil

Os mais relevantes questionamentos envolvendo o enovelamento de uma proteína e seus mecanismos surgiram na década de 60, quando Anfinsen, em seus experimentos de enovelamento sugere que a sequência de aminoácidos carregava informações suficientes para a proteína alcançar o estado tridimensional funcional da proteína (também chamado como estado nativo), sendo esse o mínimo global da energia livre (16). Esses experimentos foram realizados controlando a desnaturação da Ribonuclease Pancreática Bovina A por meio da adição de ureia ao sistema e analisando como as 4 pontes dissulfetos presentes na estrutura da proteína se comportavam. Anfinsen estabelece então uma hipótese conhecida como *"Hipótese Termodinâmica"*, assumindo que as proteínas, em seu estado nativo e condições fisiológicas normais, estão em seu estado de mínima Energia Livre.

No final da mesma década, Levinthal questionou a hipótese termodinâmica diante da inviabilidade temporal de uma busca aleatória ao longo de todo o espaço conformacional, a fim de encontrar a conformação de menor energia livre em um intervalo temporal razoável, constituindo uma busca incompatível com os requisitos biológicos (17). Um exemplo ilustrativo consiste em considerar uma proteína com cem aminoácidos em que cada aminoácido pode assumir apenas duas configurações distintas. Em termos quantitativos, para essa proteína varrer todo o espaço conformacional, assumindo essas duas conformações e assumindo que o acesso a cada conformação leva pelo menos um picosegundo (tempo de uma flutuação térmica), seria necessário então um tempo da ordem de 10^{18} segundos, que é comparável com a idade do universo. Esse dilema ficou conhecido como o Paradoxo de Levinthal (17).

Motivados pelo paradoxo proposto por Levinthal, novos estudos surgiram associando o problema do enovelamento com algumas teorias já conhecidas, como random energy model, processos de enovelamento por nucleação ou colapso hidrofóbico, spin-glass e energia de relevo (Energy Landscape) (18–21).

Introduzido no final da década de 80, a teoria de enovelamento de proteínas (18, 21), entende que o processo de enovelamento compreende uma visão global sobre a superfície de energia da proteína. Assim assume-se a ocorrência de um balanço entre entropia configuracional e energia do sistema, que direciona a proteína para o seu mínimo de energia livre, com um relevo de energia similar a um funil (22–25). Assim, o processo de

enovelamento da proteína pode ser entendido como uma reação de dois estados podendo assumir diferentes rotas nessas transições.

A teoria de funil de enovelamento assume uma superfície de energia com característica multidimensional, que permite por seus fundamentos analisar o processo de enovelamento protéico por meio de uma ou mais coordenadas de reação. Na Figura 2 é apresentada uma representação de superfície de energia afunilada, onde o topo do funil é representado pelas diversas estruturas possíveis da proteína em seu ensemble desenovelado. Ao longo do corpo do funil, o número de conjuntos de estruturas possíveis vai se reduzindo. Esse processo ocorre até que a proteína alcance o fundo do funil, sendo esse o que chamamos de estado nativo, ou estado de mais baixa energia, que é não-degenerado. Essa nova perspectiva do enovelamento de proteínas nos indica que pode haver um ou mais caminhos no processo de enovelamento guiados pelo balanço energia-entropia do sistema, permitindo inclusive uma multiplicidade de possíveis rotas (22).

A energia livre F(Q) é um potencial termodinâmico escolhido para descrever o balanço da contribuição energética E(Q) e a contribuição entrópica S(Q) do sistema:

$$F(Q) = E(Q) - TS(Q) \tag{1}$$

A energia livre de Helmholtz é utilizada aqui devido ao emprego do ensemble canônico (NVT) nos estudos. F(Q) é o parâmetro que descreve a evolução do sistema em função de uma coordenada de reação Q.

A teoria de funil de enovelamento é empregada neste trabalho na modelagem dos seguintes sistemas e explorando dois temas distintos:

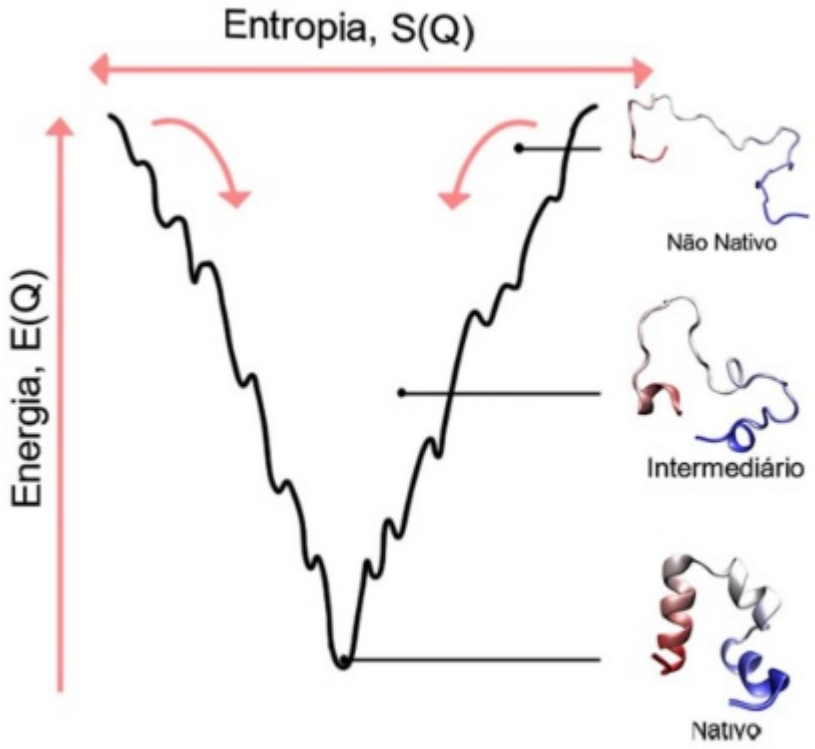
I) a proteína GRB2, uma proteína adaptadora com multidomínios, onde são avaliados os mecanismos que se relacionam ao seu enovelamento. Essa é uma proteína de interesse devido à sua organização estrutural em três diferentes domínios que se contrastam aos processos biológicos em que ela aparece envolvida, o que será melhor detalhado no capítulo 1;

II) em relação às MPTs, em especial as glicosilações, com a finalidade de compreender os efeitos das glicosilações ao longo do processo de enovelamento de proteínas, tema que ainda é fonte de muitas especulações. Assim, é feito um estudo comparativo de três proteínas partindo de suas estruturas resolvidas no PDB. Os carboidratos (glicosilações) estão presentes nas estruturas resolvidas e são modelados para os sistemas conforme será

apresentado na seção de métodos específicos.

Na próxima seção estão descritos os métodos gerais empregados e nos capítulos subsequentes os resultados com detalhamento de métodos específicos quando estes se fazem necessários.

Figura 2: Representação do funil de energia. Ilustração do funil de estruturas do processo de enovelamento de uma proteína ao longo da coordenada de reação Q. Na figura são apresentadas as componentes Entropia S(Q) e Energia E(Q). As estruturas desenhadas à direita representam os respectivos ensembles das condições nomeadas abaixo com um ponto indicativo da conformação no funil.



Fonte: Adaptado de (*https://en.wikipedia.org/wiki/Folding-funnel*) (2023).

2 Metodologia

2.1 Modelos baseados em estrutura

A complexibilidade das proteínas exige um tratamento adequado para o entendimento de seu funcionamento e de processos biológicos em que elas estejam envolvidas, por vezes sendo as melhores estratégias aquelas associadas à simplicidade. A utilização de modelos mais realísticos envolve uma infinidade de parâmetros e de interações atômicas, e esses parâmetros muitas vezes não atendem aos recursos de infraestrutura disponíveis ou ainda se restringem ao entorno de um mínimo local energético específico. A utilização de modelos reduzidos para dinâmicas moleculares pode ser uma estratégia eficiente para o estudo e entendimento do enovelamento de proteínas (26–30). Esses modelos permitem que as simulações varram extensivamente o espaço conformacional com razoável amostragem, possibilitando os cálculos de propriedades termodinâmicas de interesse por meio de fundamentos da física estatística.

Em sua construção mais simples, os resíduos de aminoácidos são representados por esferas localizadas nas posições dos carbonos alfas, atribuindo a eles um único raio de 4Å. O potencial do modelo baseado na estrutura é construído a partir da conformação nativa da proteína, o que dá origem ao seu nome, todavia o mesmo modelo também é referenciado como Gö-like (31, 32). Essas conformações do estado nativo podem ser obtidas por técnicas experimentais como cristalografia de raio-X e ressonância magnética nuclear (RMN) ou também por modelagens teóricas por homologia ou inteligência artificial (33, 34). A energia potencial de uma dada configuração Γ da proteína em relação a sua configuração no estado nativo Γ_0, é dado pela expressão:

$$\begin{aligned} V(\Gamma,\Gamma_o) &= \sum_{bonds} \epsilon_r (r-r_o)^2 + \sum_{angles} \epsilon_\theta (\theta-\theta_o)^2 \\ &+ \sum_{backbone} \epsilon_\phi \left\{ [1-\cos(\phi-\phi_o)] + \frac{1}{2}[1-\cos\left(3(\phi-\phi_o)\right)] \right\} \\ &+ \sum_{contacts} \epsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \end{aligned} \quad (2)$$

Na equação, todas as variáveis apresentadas com o índice 0 indicam os valores obtidos

em sua estrutura nativa (ou estrutura de referência dependendo do contexto). Os pré-fatores energéticos são: ϵ_r (ligação covalente), ϵ_θ (angulares), ϵ_ϕ (diedrais), ϵ_C (Lennard Jones) e ϵ_{NC} (repulsão). Seus valores são apresentados em unidades reduzidas em relação à energia de interação entre contatos que é dado por um valor unitário, assim $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$ e $\epsilon_\phi = \epsilon_0 = \epsilon_C = \epsilon_{NC}$, onde ϵ_0 tem o valor de 1 kcal/mol (34). Ainda na equação, r representa a distância entre dois átomos ligados covalentemente, θ é o ângulo entre três átomos consecutivos da estrutura nativa e, analogamente, ϕ é o ângulo diedral formado entre quatro átomos consecutivos, especificamente para o modelo Cα. O parâmetro d_{ij} do termo de Lennard-Jones é determinado através do mapa de contatos entre os átomos na estrutura nativa da proteína para os resíduos "i" e "j" em questão. O mapa de contatos é obtido utilizando o algoritmo *Contact of Structural Units* (CSU) (35). O parâmetro σ_{NC} compõe o termo repulsivo para ligações não contempladas no mapa de contato e garante as restrições estéricas entre os resíduos de aminoácidos. Nesse caso, essa distância é maior que $\sigma_{NC} = 4$Å. Esses parâmetros são obtidos a partir da estrutura nativa alvo para o estudo, e definem o mínimo global da energia potencial. Os arquivos de entrada para a simulação podem ser obtidos utilizando a interface gráfica do servidor SMOG (36)

2.2 Coordenadas de Reação

As coordenadas de reação, também conhecidas como parâmetros de ordem, são variáveis utilizadas para acompanhar a evolução do enovelamento protéico em uma simulação computacional (ou em experimentos no geral). A coordenada de reação, se adequadamente escolhida, deve possibilitar a descrição do processo de enovelamento e suas nuances. Um parâmetro de ordem recorrente em estudos de enovelamento de proteínas é o número de contatos nativos (Q). Esta coordenada de reação utiliza de uma conformação de referência com um número total de contatos calculados a partir de algum método, em nosso caso o CSU (35), para calcular a quantidade de contatos formados numa determinada conformação diferente da referência.

Uma outra coordenada de reação frequentemente utilizada é o RMSD (Root Mean Square Deviation) que mede o desvio médio quadrado da distância entre os resíduos da proteína com base em uma estrutura de referência ao longo da simulação. Muitas outras coordenadas de reação podem ser empregadas dependendo do estudo de interesse como por exemplo o Raio de Giro (Rg), distância entre os resíduos terminais da cadeia (distância *end-to-end*), distância específica entre dois resíduos ou do centro de massa de dois domínios,

entre outras. Todas essas coordenadas tem suas particularidades e devem ser utilizadas de acordo com os objetivos a serem alcançados para a análise com a avaliação crítica do executor. Até então não temos um parâmetro universal que permita uma descrição completa de todo o mecanismo de enovelamento pela própria complexidade do processo de enovelamento, assim como suas características locais e globais.

2.3 Valores *phi*

Valores *phi* (na linguagem padrão da literatura internacional, ϕ-*values*) (37) é uma metodologia para avaliar informações estruturais dos estados de transição ao longo do enovelamento de proteínas, indicando resíduos de aminoácidos essenciais para o processo no estado de transição. Essa região é definida como o intervalo delimitado pelos valores de Q (fração de contatos nativos) que cortam o grafico de energia livre por uma reta a $1K_BT$ abaixo do ponto máximo. Experimentalmente esse método avalia todas as possíveis mutações e quantifica as contribuições de cada resíduo mutado (38).

$$\Phi = \frac{\Delta G^{TSE-Desenovelada}_{Nativa} - \Delta G^{TSE-Desenovelada}_{Mutada}}{\Delta G^{Nativo-Desenovelada}_{Nativa} - \Delta G^{Nativo-Desenovelada}_{Mutada}} \tag{3}$$

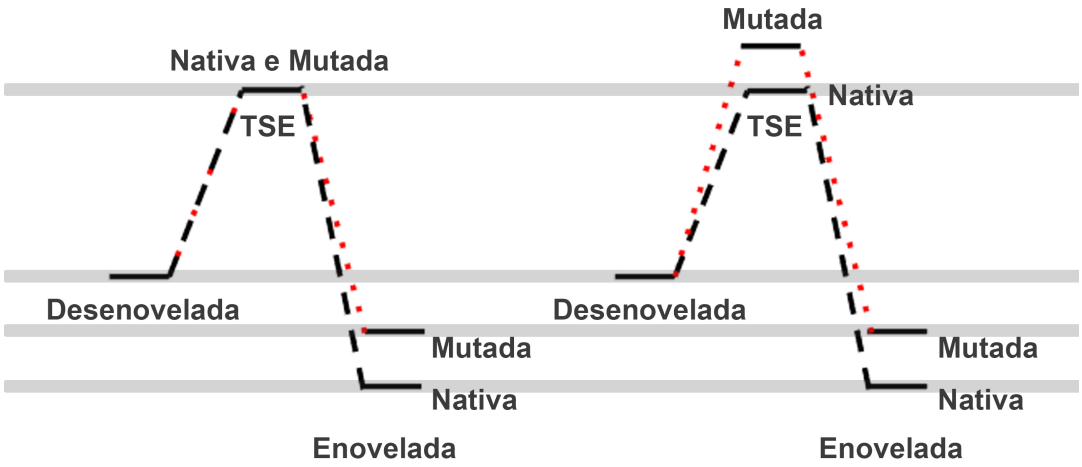
onde: $\Delta G \approx \ln < e^{\frac{\Delta E}{K_B T}} >$

$$\Phi = \frac{\Delta\Delta G^{TSE-Desenovelada}}{\Delta\Delta G^{Nativo-Desenovelada}} \tag{4}$$

para uma proteína de dois estados, ΔG significa as diferenças de energia livre entre os estados desenovelados e de transição e entre os estados desenovelados e estados enovelados respectivamente. O primeiro Δ do "$\Delta\Delta$" denota diferenças de energia livre provocadas por uma mutação pontual no resíduo de aminoácido alvo do cálculo. Os valores de ϕ são calculados para todos os resíduos de aminoácidos da proteína. Valores próximo de 1 indicam que o resíduo em questão tem ligações formadas no estado de transição, e assim ele é considerado importante para que o enovelamento ocorra, enquanto um valor ϕ próximo

de 0 significa que o resíduo não está envolvido na estabilização do estado de transição, como mostrado na figura 3.

Figura 3: Diagrama do Estado de Transição (TSE). O diagrama ilustra a energia dos estados desenovelados, de transicão e enovelados. Em casos onde o resíduo mutado não interfere na formação de TSE, os valores de ϕ são próximos de 0 e representados pelo esquema da esquerda. Quando o resíduos mutado interfere na formação de TSE, ϕ é próximo de 1 e o esquema que representa essa situação está a direita do diagrama.



Fonte: Adaptada de (39) (2023).

Uma segunda forma de calcular os valores Φ nas simulações e que também será utilizado neste trabalho, é utilizar a probabilidade de formação dos contatos ao longo de janelas da nossa coordenada de reação. Assim é possível verificar a probabilidade do contato ser formado em um determinado range da coordenada utilizada. A equação para esse calculo é dada por:

$$\Phi_i = \frac{P^i_{TSE} - P^i_{Desenovelada}}{P^i_{Enovelada} - P^i_{Desenovelada}} \tag{5}$$

sendo P_i a probabilidade de formar um contato i no estado proposto. Valores de $\phi = 0$ indicam que o contato nunca é formado no TSE e portanto não são importantes para a

proteína vencer a barreira de energia livre. Ao contrário, $\phi = 1$ indica que esse contato é fundamental para que a proteína alcance o estado nativo.

2.4 Rotas de Enovelamento

O cálculo de rotas de enovelamento de uma proteína nos permite quantificar a multiplicidade de caminhos que a proteína pode assumir ao longo do processo de enovelamento. Deste modo, é possível avaliar se a proteína enovela por um ou mais conjuntos de conformações sequencialmente definido (*pathways*). Ambas situações podem ocorrer em proteínas ou em trechos específicos ao longo da coordenada de reação, sendo o alvo desse tipo de análise.

O cálculo da rota de enovelamento fornece informações sobre a especificidade na formação estrutural das proteínas, relacionando-se inclusive com a entropia configuracional do enovelamento (40, 41). Em outras palavras, calculamos quão específica é a formação de uma determinada estrutura ao longo do enovelamento utilizando a nossa coordenada de reação como o parâmetro quantitativo desse processo. A equação que define o cálculo da rota de enovelamento R(Q) é dada pela seguinte expressão:

$$R(Q) \quad = \quad \sum_{i=1}^{M} \frac{(< Q_i > -Q)^2}{MQ(1-Q)} \tag{6}$$

Em que Q_i especifica a formação de contato i em uma estrutura particular, e a média é tomada sobre todas as configurações Q com contatos formados do total M. O valor de R(Q) varia entre 0 e 1, sendo que quanto mais específico for o caminho de enovelamento da proteína esse valor é mais próximo de 1, de tal forma que, podemos idealizar um valor de R(Q)=1 em que proteína sempre seguiria pela mesma rota para se enovelar. Por outro lado, um valor de R(Q) próximo de zero significa que qualquer contato nativo possui a mesma probabilidade de ser feito, ou seja, neste caso, a proteína não teria rotas específicas.

2.5 Energy Landscape Visualization Method (ELViM).

O software ELViM é uma alternativa para problemas de enovelamento de proteínas em que as coordenadas de reação não conseguem descrever de maneira única e sem sobreposição

de valores nos estados de transição. É um método que utiliza uma métrica baseada na distância interna entre aminoácidos para descrever as diferenças entre duas conformações (42, 43). No primeiro passo, o método define uma métrica utilizando as posições dos aminoácidos, cujas coordenadas internas caracterizam as diferentes conformações. De maneira formal, a medida de comparação entre duas conformações (k, l) pode ser escrita como:

$$q^{k,l} = \frac{1}{N} \sum_{N} \exp \frac{-(r_{i,j}^{k} - r_{i,j}^{l})^2}{\sigma_{i,j}^2}, \tag{7}$$

onde N_p é o número total de pares de resíduos, $r_{i,j}^{k(l)}$ é a distância entre o resíduo i e o resíduos j na conformação $k(l)$ e i, j é uma constante de pesagem igual a $\sigma_{i,j} = \sigma_0 |i - j|^{\epsilon}$, com $\sigma_0 = 1$ $\epsilon = 0.15$. Essa medida é calculada para todos os pares de estruturas k e l de uma dada trajetória, resultando em uma matriz simétrica $n \times n$, onde n é o número total de conformações analisadas.

O segundo passo consiste em visualizar a superfície de energia, e para isso ocorre uma redução multidimensional dos dados originais, reduzindo o espaço multidimensional original dos dados a um espaço bi e tridimensional (42–45). A técnica utilizada consiste em colocar os objetos da matriz em posições iniciais aleatórias em um espaço 2D e aplicar um passo de minimização para preservar as distâncias na matriz de dissimilaridade. O processo cessa quando a diferença nas distâncias entre as iterações consecutivas estiver abaixo de um determinado corte. No resultado final da projeção, a proximidade dos pontos indica a semelhança entre as estruturas representadas. Pontos próximos indicam instâncias semelhantes de acordo com a medida de distância δ. Intuitivamente, pontos distantes representam objetos com pouca relação, também de acordo com δ.

Todos as etapas de utilização do software ELViM foram realizados com o auxilio do aluno Murilo Sanches do grupo computacional do professor Vitor Leite.

2.6 Simulações Computacionais

As informações estruturais das proteínas estudadas ao longo deste trabalho foram extraídas do banco de dados PDB (5), assim para todos os casos serão fornecidos os seus códigos de acesso (PDB id). O software GROMACS (46), versão 4.3.4 foi utilizado para todas as simulações de dinâmicas moleculares deste trabalho. Os arquivos de topologia

(arquivos de entrada) foram gerados por meio da ferramenta online SMOG v 1.2.3 (36). As simulações foram realizadas utilizando um integrador estocástico e o acoplamento térmico de Berendsen (47). Ao todo foram realizadas simulações com tempos totais de 600ns, utilizando passos de integração dt de 0.0005ps. Foram utilizadas para todas as simulações caixas cúbicas, com condições periódicas de contorno e suas bordas distanciando 10Åda superfície da proteína. Para cada uma das proteínas simuladas, o intervalo de temperaturas é ajustado de acordo com as respostas de cada sistema, i.e, primeiramente analisando uma gama de temperaturas igualmente espaçadas e depois restringindo-se as temperaturas de interesse e suficientes para uma adequada estatística.

2.7 Trabalhos adicionais e complementares

Os métodos descritos nas seções anteriores são referentes ao principal tópico deste projeto de doutorado: análises de enovelamento utilizando modelo baseado em estrutura. Além dessa abordagem, trabalhos em paralelo também foram realizados abordando análises de docking, interação proteína-ligante, diferenças de pH e complexação de copolímeros com estruturas de amilopectina. Estes trabalhos não serão descritos detalhadamente na tese final, todavia podem ser encontrados nos anexos como trabalhos publicados ou submetidos.

CAPÍTULO I

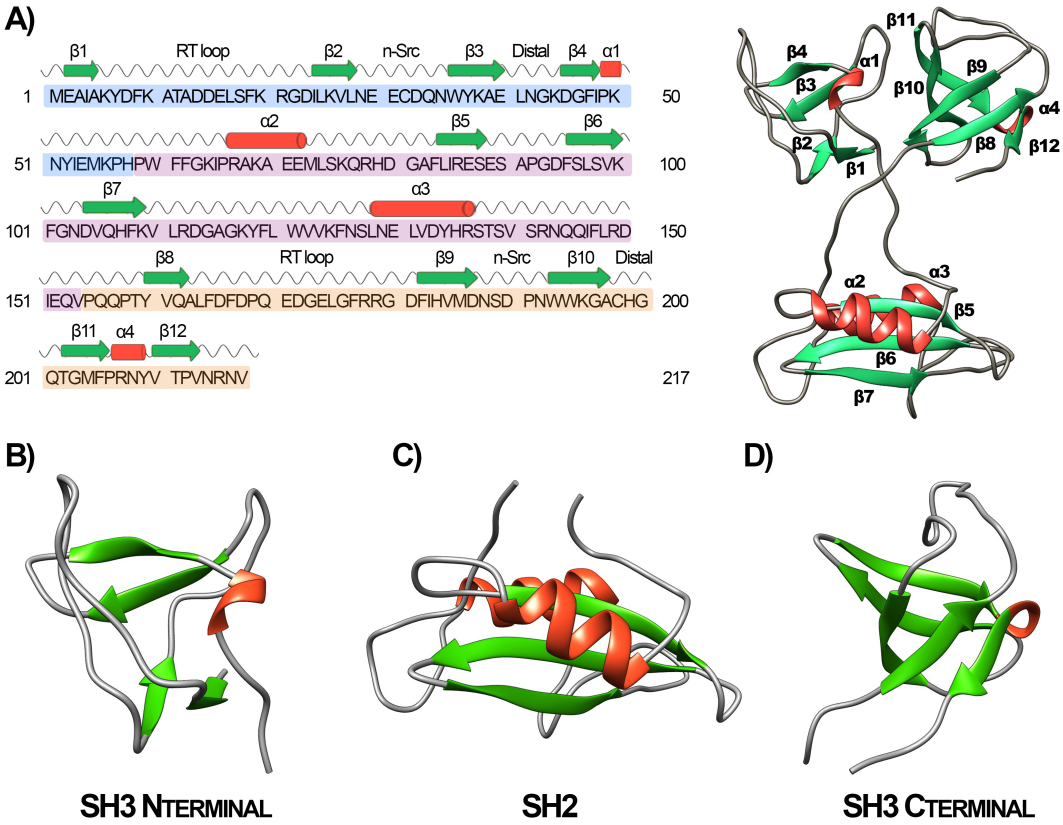
Growth Factor Receptor Bound 2 (GRB2)

3 *Growth Factor Receptor Bound 2 (GRB2)*

3.1 Introdução

A proteína GRB2 (do inglês, *Growth Factor Receptor Bound 2*) é uma proteína adaptadora citosólica, ubiquamente expressa em humanos. Composta por 217 resíduos de aminoácidos e com uma massa molecular de 25 kDa, é estruturalmente organizada por um domínio SH2 flanqueado pelos domínios SH3s (N e C terminal), como mostrado na figura 4.

Figura 4: Proteína GRB2. A) Sequência de aminoácidos da proteína GRB2 monomérica, as sequências são coloridas de acordo com o domínio pertencente: SH3 $N_{TERMINAL}$ em azul, SH2 em roxo e SH3 $C_{TERMINAL}$ em laranja. Em destaque também são mostradas as estruturas secundárias seguindo a construção da estrutura 3D. B, C e D) Domínio SH3 $N_{TERMINAL}$, SH2 e SH3 $C_{TERMINAL}$ respectivamente. Todos eles são coloridos de acordo com as estruturas secundárias, sendo em verde as folhas betas e laranja as hélices alfas.



Fonte: Elaborada pelo autor (2022).

Dentre os domínios que compõem a GRB2, o domínio SH2 (*Src homology-2*) é composto por uma estrutura constituída de aproximadamente 100 resíduos que se enovelam em

uma folha-β antiparalela constituída por cinco fitas (Figura 4C) (48, 49). Essa folha-β encontra-se no centro de SH2 e divide o domínio em dois sítios com funções diferentes. O sítio I contém três resíduos de aminoácidos carregados positivamente (ARG67, ARG86 e LYS109), cujas cadeias laterais são responsáveis pelo reconhecimento da fosfotirosina (pY) de outras proteínas, em contrapartida, o sítio II é um bolso hidrofóbico responsável por reconhecer os peptídeos $pY + 1$ a $pY + 3$ em proteínas parceiras (50, 51).

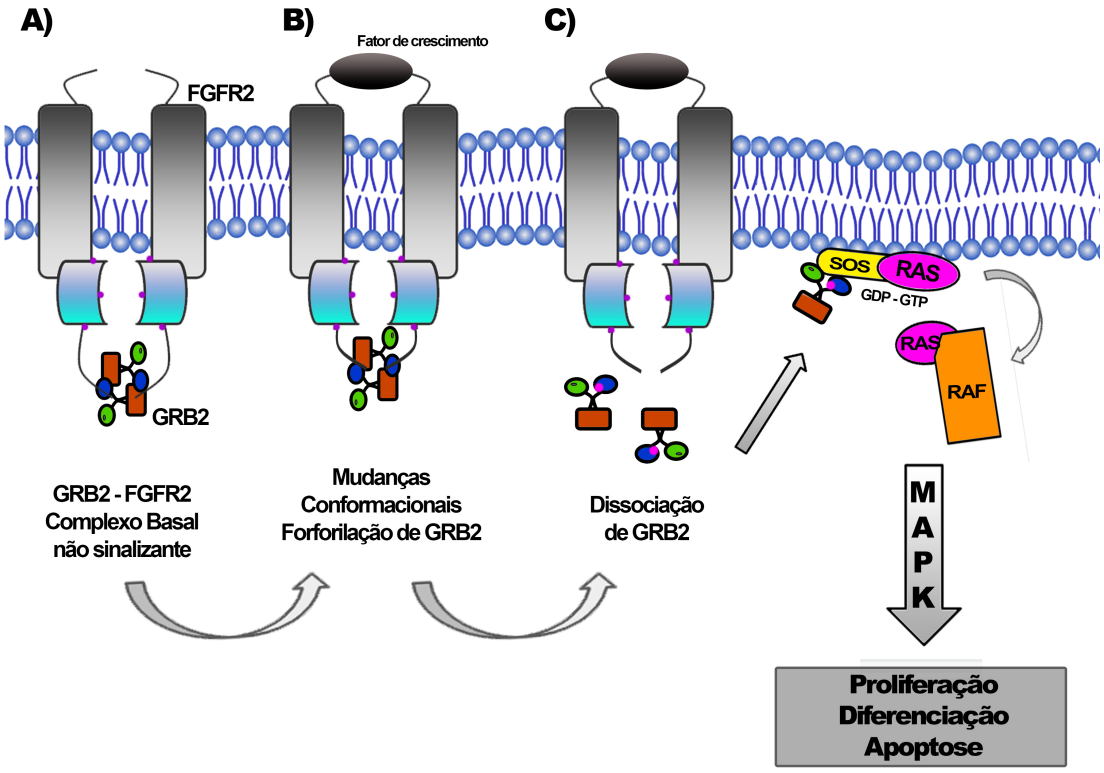
Os domínios SH3s (Figura 4B e D) são estruturas encontradas em mais de 220 proteínas diferentes sendo responsáveis por diversas funções celulares fundamentais, como regulação, proliferação e migração (52) Esses domínios são estruturalmente formados por uma pequena hélice 3-10 e de quatro a oito fitasβ em um arranjo de duas folhas-β ou barril β. As alças entre as folhas são denominadas RT, n-Src e distal, sendo regiões essenciais no reconhecimento de outras proteínas que interagem com GRB2 . Outros estudos revelaram inicialmente que o domínio SH3 tem afinidade para interação em motivos PxxP, acomodando-o em dois sítios diferentes (53, 54). Além disso, peptídeos contendo o motivo PxxP poderiam interagir através de duas conformações opostas, definidas pela posição relativa de um resíduo carregado positivamente (+xxPxxP ou xPxxPx+). Tais conformações são essenciais em como o domínio interage com outras proteínas (52–56).

Em sua forma completa, a GRB2 possui um arranjo tridimensional estrategicamente elaborado pela natureza, de forma que cada um de seus domínios tem liberdade para realizar mudanças conformacionais sem alterar as outras regiões, e consequentemente facilitar a interação com outras proteínas ou até mesmo pequenas moléculas. Toda essa plasticidade e possibilidades faz com que GRB2 seja uma proteína essencial para a transdução de sinais intracelulares, enquanto que suas especificidades de interações fazem com que a GRB2 atue como 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 (57).

Um exemplo sobre sua importância na célula é a sua interação com a proteína quinase FGFR2, uma importante proteína transmembrana responsável pela ativação da via de sinalização MAPK (58, 59). Essa interação é intermediada pela falta de estímulos extracelulares de fatores de crescimento, formando um complexo GRB2-FGFR2 denominado "nível de fosforilação basal", um estado fosforilado e não sinalizante. A ligação de fatores de crescimento na porção extracelular de FGFR2, provoca mudanças conformacionais em sua parte intracelular e a fosforilação de alguns resíduo de tirosina dentro da estrutura de

GRB2, dissociando o complexo via repulsão eletrostática (59). Ressalta-se aqui que todo o processo de interação GRB2-FGFR2 acontece com GRB2 estruturada em dímero, e após a fosforilação de seu resíduo e a dissociação, GRB2 está estruturada em monômero, como mostrado pela figura 5.

Figura 5: Representação do mecanismo de regulação da proteína quinase FGFR2 através da proteína adaptadora GRB2. A proteína FGFR2 é representada mimeticamente pelas barras cinzas simbolizando sua região citoplasmática e com seu domínio quinase 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 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 sinalização..



Fonte: Adaptada de (59) (2022).

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 (59). 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 (60).

Como mostrado, o constante equilíbrio entre as formas dimérica e monomérica de GRB2 também é um fator importante para a regulação da via de sinalização, sendo cada uma delas associada a diferentes funções (61). Essa complexibilidade de fatores promove a proteína GRB2 da função apenas de adaptadora para um importante regulador de processos e vias metabólicas, funcionando como um mecanismo semelhante a um interruptor, onde apenas seu estado monomérico é sinalizador competente para ativar a via de sinalização MAPK, enquanto o estado dimérico é a sua forma inibitória.

Toda essa complexibilidade dos processos envolvidos com GRB2 tem como ponto principal o seu rearranjo estrutural, levando ao interesse principal desse trabalho que é analisar computacionalmente a dinâmica de seus domínios, considerando a proteína GRB2 em sua forma monomérica e também seus domínios isolados. Verificar seu enovelamento, estados intermediários, flutuações responder diversas questões dos mecanismos básicos por trás desses processos além de possibilitar futuros estudos de construção de ligantes específicos.

A colaboração entre resultados teóricos e experimentais para estudos de interações da GRB2 com moléculas de interesse farmacêutico que poderiam intervir em processos de sinalização celular descontrolada e consequentemente em possíveis desenvolvimentos tumorais foi realizado e encontram-se em anexo. Esses resultados foram pioneiros para novos trabalhos (62) e pesquisas ainda em andamento pelo grupo do Prof. Dr. Fernando Melo da Unesp de São José do Rio Preto (IBILCE), envolvendo os domínios de Grb2 separados.

3.2 Objetivos

Estudar o enovelamento da proteína GRB2 utilizando das informações já consolidadas sobre suas estruturas e mudanças estruturais. Relacionar os resultados teóricos computacionais com os resultados de experimentação para avançar o entendimento sobre os mecanismos que envolvem a proteína GRB2. Em paralelo a esses estudos, utilizaremos das

informações experimentais em colaboração com o Laboratório de Biofísica Molecular, do Prof. Dr. Fernando Alves de Melo para experimentos de interação com pequenos ligantes. Esses últimos objetivos têm seus métodos, resultados e conclusões detalhados nos anexos em forma de artigo.

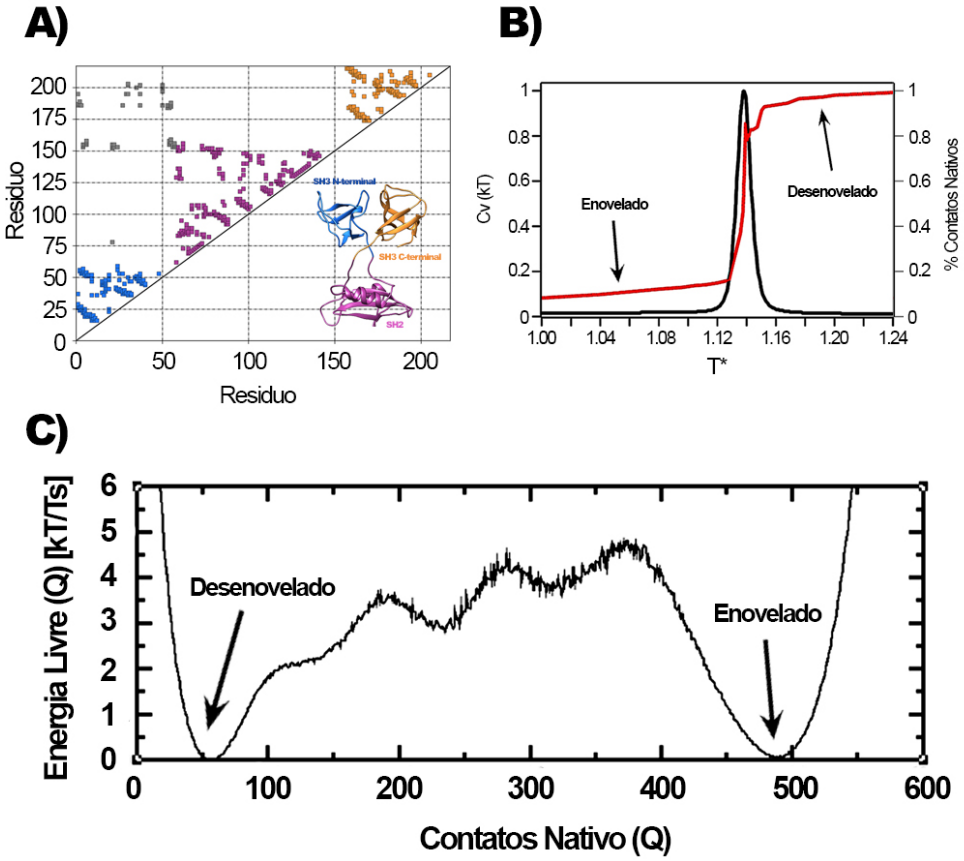
3.3 Resultados e Discussões

3.3.1 Perfil de Enovelamento da proteína Grb2

Como ponto de partida para os estudos de enovelamento da proteína GRB2, foram realizadas dinâmicas moleculares utilizando o modelo baseado em estrutura $C\alpha$. O mapa de contato é calculado com base em sua estrutura cristalográfica, incluindo os contatos interdomínios, possibilitando análises de fenômenos dentro de cada domínio isolado assim como a intercomunicação proveniente dos contatos entre eles. Esses contatos podem ser vistos na figura 6 onde cada ponto representa o contato do resíduo em X com o resíduo em Y, além disso, a imagem é colorida de acordo com cada domínio da GRB2 mostrados na estrutura 3D.

Dentro desse primeiro conjunto de resultados, calculamos o calor específico a volume constante (Cv) e a porcentagem de contatos nativos (Figura 6C) assim como o perfil de energia livre (Figura 6C). Todos esses resultados em conjunto nos mostram que existe a transição enovelado/desenovelado e que durante o enovelamento a GRB2 passa por diversos estados de mínimos de energia, sendo que a última transição para chegar até o estado totalmente enovelado é energeticamente maior em relação as outras transições.

Figura 6: Perfil termodinâmico da proteína GRB2. A) Mapa de contatos calculados utilizando o algoritmo CSU. Cada ponto no gráfico representa o contato do resíduo em X com o resíduo em Y. Os contatos são coloridos de acordo com o domínio onde é encontrado, sendo azul SH3 NTERMINAL, Roxo SH2 e Laranja SH3 CTERMINAL. Em Cinza também são representados os contatos entre os domínios. B) Calor Específico a volume constante e fração de contatos, mostrando a temperatura de transição (dado em unidades reduzidas). C) Energia Livre em função da coordenada de reação dos contatos nativos (Q). A curva mostra diversos mínimos de energia entre os estados enovelados e desenovelados, mostrando um enovelamento que passa por estados intermediários. O valor da barreira de energia também mostra que a última transição (intermediario/Nativo) é a mais custosa energeticamente em relação as outras transições

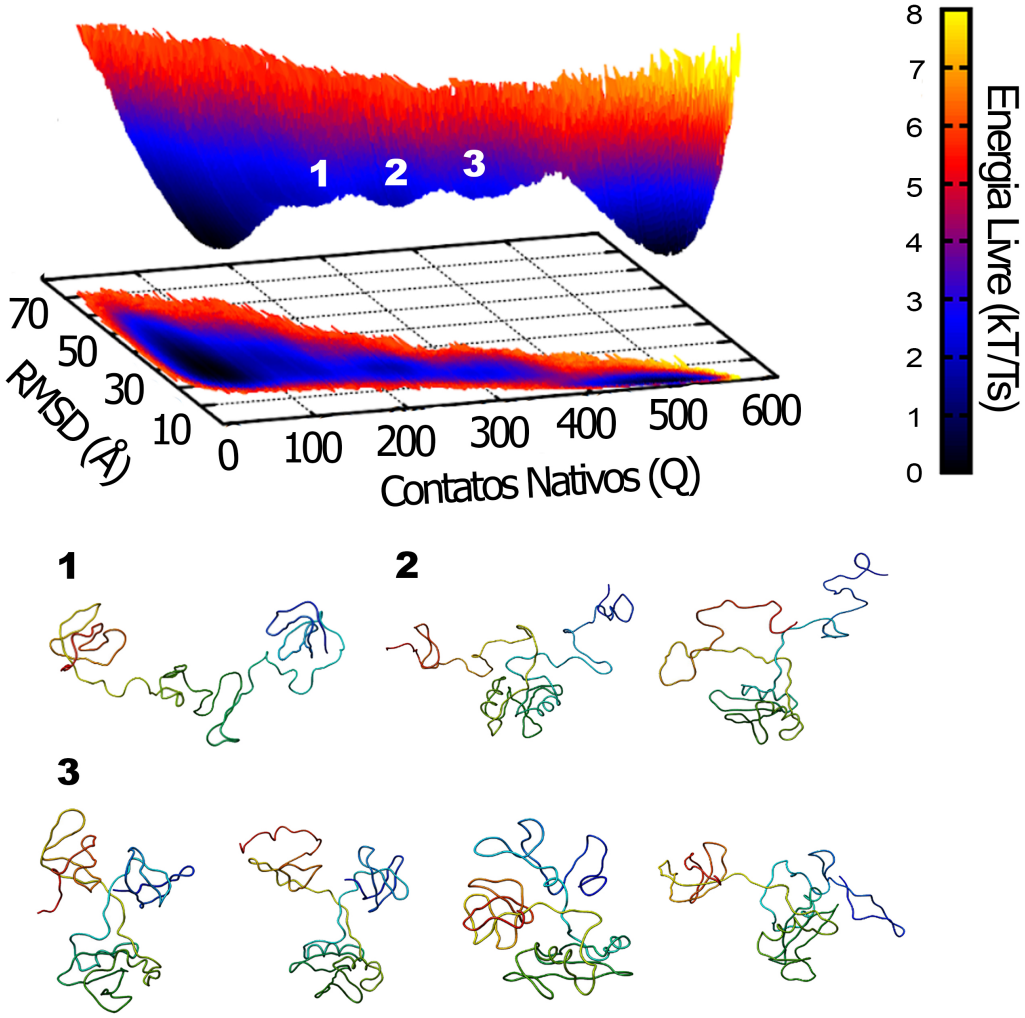


Fonte: Elaborada pelo autor (2023).

Para descrever esses estados, optamos por utilizar uma segunda coordenada de reação, no caso, o RMSD, e construímos a superfície de energia, correlacionando as duas coordenadas nos eixos XY e colorindo em relação a Energia Livre (F(Q)). Essa correlação nos dá segurança de ter um resultado de duas diferentes perspectivas, além de nos permitir discutir acerca de suas estruturas formadas em cada um desses poços de mínima energia, uma vez que a coordenada RMSD é um parâmetro físico estrutural. Para essa análise, retiramos todas as estruturas presentes em cada um dos mínimos e clusterizamos em pequenos

conjuntos, como mostrado na figura 7.

Figura 7: Superfície de energia e clusterização de estruturas. A) Superficie de energia livre construída utilizando a correlação das coordenadas de reação dos contatos nativos (Q) e RMSD. Essa superfície é colorida de acordo com a energia livre de cada estado, mostrando mais uma vez 3 estados intermediários entre os estados nativo e desenovelado. Os números destacados (1,2 e 3) foram utilizados para a clusterização das estruturas pertencentes aqueles estados, sendo representados via tubos em 3D, coloridos utilizando a paleta "*rainbow*" do vermelho para o azul.



Fonte: Elaborada pelo autor (2022).

Esse resultado nos indica um novo entendimento do que estava acontecendo em nosso sistema. Uma vez que a estrutura de GRB2 é formada por três diferentes domínios, esses estados de mínima energia se referem a estados parcialmente enovelados. Em outras palavras, o enovelamento desses domínios não são concomitantes, e nossa hipotése em relação a esses primeiros resultados era de que primeiro se forma um deles, acreditando-se

nesse caso ser o SH2 para depois formar-se o segundo e finalizar formando o terceiro, sendo quaisquer um dos SH3s nessas ultims formações.

3.3.2 Formação de contatos e especificidade do enovelamento

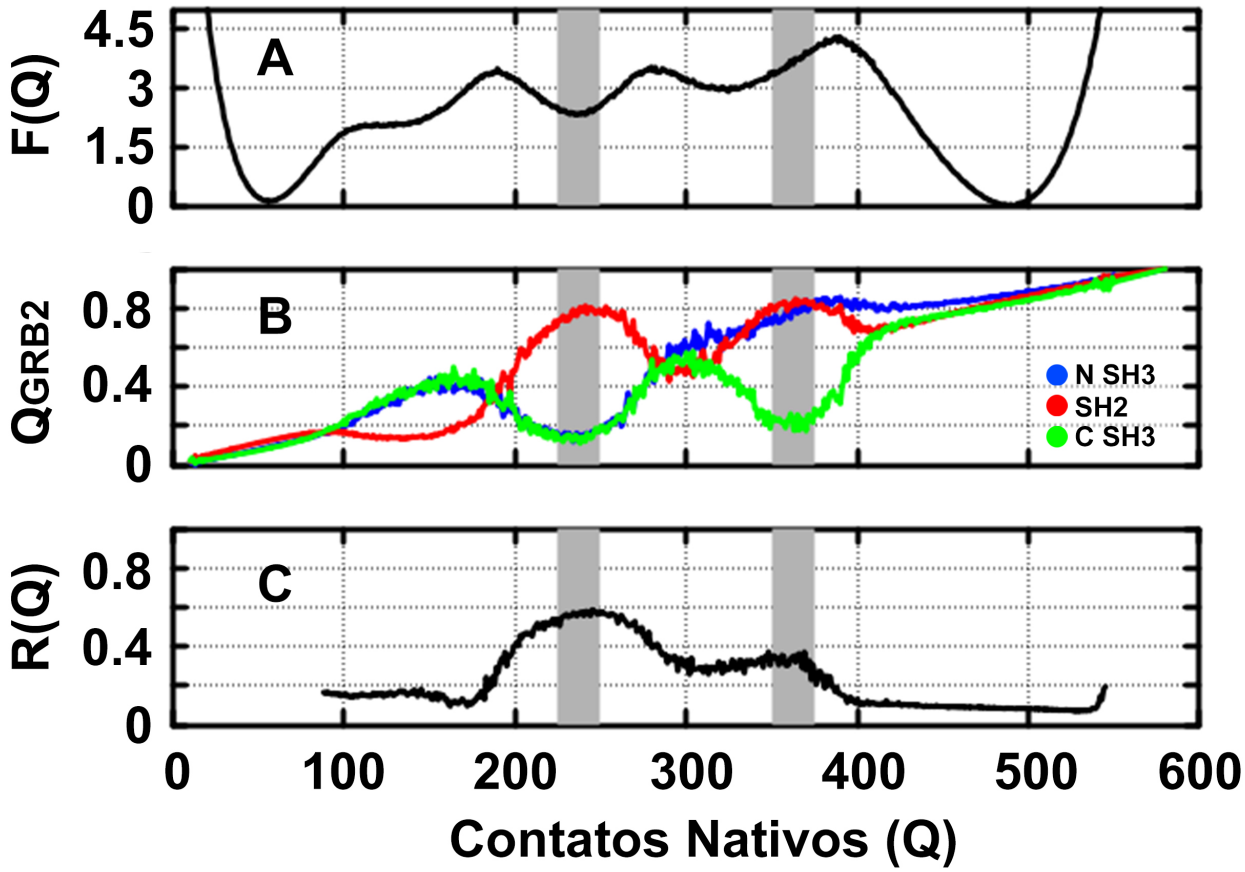
Uma vez que o enovelamento de GRB2 não é concomitante entre seus domínios, formando-se um de cada vez e possibilitando alguns estados semi-enovelados, partimos para as análises de como estes estão sendo formados. Calculando a rota de enovelamento R(Q), conseguimos ver a especificidade da formação de algumas estruturas de acordo com a coordenada de reação Q, como mostrado na Figura 8C. Uma segunda análise foi separar os contatos referentes a cada domínio e acompanhar a porcentagem desses formados ao longo da nossa coordenada de reação, em outras palavras, aqui conseguimos distinguir quando está sendo formado um domínio ou outro como mostrado na Figura 8B.

A rota de enovelamento mostrou dois picos de de R(Q) com valores próximos a 0.6 e a 0.4, indicando a formação de 2 estruturas específicas e que quando sobrepostas com a formação dos contatos, a coordenada de reação mostra que essa especificidade de formação é em relação aos domínios SH2s e SH3 N-terminal.

A ordem dos eventos mostrados na figura 8 seguindo a formação dos domínios, nos levaram a hipóteses de estar ocorrendo fenômenos como backtracking, começando por exemplo com a formação de alguns contatos nos domínios SH3C e SH3N terminal seguido pela formação dos contatos do domínio SH2 em troca da quebra de alguns contatos que já haviam sido formados. Esse processo incide com o primeiro pico calculado na rota de enovelamento, indicando ser um processo recorrente na formação dos primeiros contatos de GRB2.

Todavia, essas hipóteses precisavam de maiores descrições para serem formalizadas. O principal problema encontrado nessas análises é o fato da proteína GRB2 ser uma proteína multidominios com estados semi-enovelados. Ao separarmos a quantidade de contatos em cada domínio, há uma sobreposição de valores seguindo a coordenada de reação Q, isto é, como GRB2 é formado por dois domínios SH3s, o número de contatos totais que moldam esses domínios têm valores bem próximos, ainda, o SH2 que é um domínio maior que os SH3s tem um número de contatos totais próximos a soma dos contato de SH3C e SH3N. Essa sobreposição ocasionou na necessidade de um segundo ponto de vista de como está sendo todo esse processo de enovelamento, então utilizamos o software ELViM (43) o qual

Figura 8: Rota de enovelamento e Formação de contatos ao longo do enovelamento. (A) Perfil de Energia Livre. (B) Formação dos contatos separados por domínios. Cada cor representa os contatos dentro de um determinado domínio (vermelho - SH3N, preto - SH2 e verde SH3C), as porcentagens calculadas mostram quanto de cada um desses domínios estão formados de acordo com a coordenada de reação dos contatos nativos. (C) Rota de enovelamento R(Q) que é calculada acompanhando a coordenada de reação Q (contatos nativos) ao longo da dinâmica. Os dois picos mostram que nesses pontos há uma recorrência na formação de uma determinada estrutura.



Fonte: Elaborada pelo autor (2022).

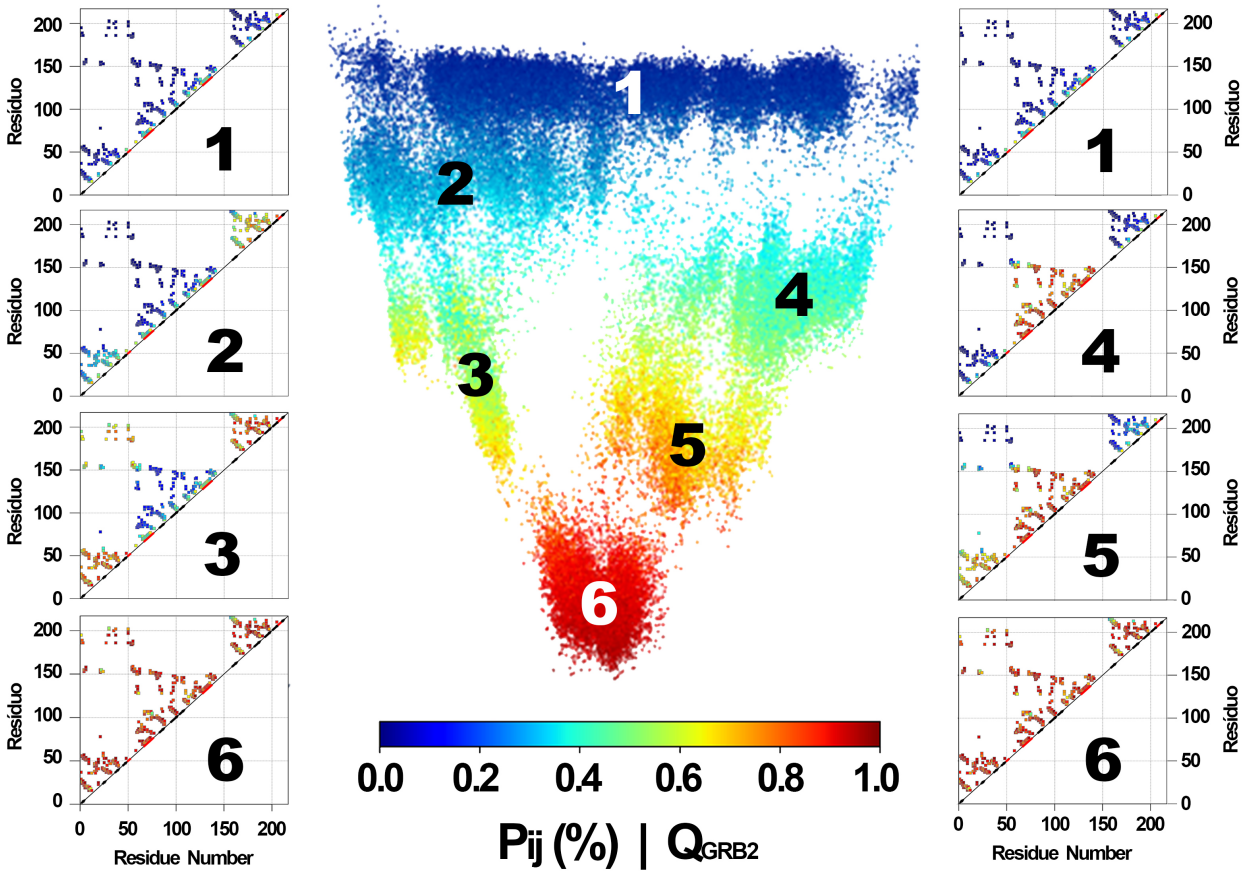
já tem incluso em seus cálculos a coordenada de reação (q) possibilitando uma comparação direta com as coordenadas utilizadas até então.

3.3.3 Funil de energia e estados semi-enovelados

Os resultados obtidos via ELViM elucidaram uma nova perspectiva. É sugerido computacionalmente que o enovelamento de GRB2 não segue uma construção específica e contínua dos seus domínios e sim a possibilidade dois possíveis caminhos partindo do estado desenovelado até se enovelar por completo. A projeção 3D mostrada na Figura 9

tem a forma de um funil, com as conformações desenoveladas na região superior (em azul escuro) passando pelos clusters referentes aos domínios formados e chegando ao ensemble enovelado na parte inferior (em vermelho). Cada ponto nessa representação corresponde a uma conformação retirada da dinâmica molecular com o eixo z correspondendo aos contatos nativos de GRB2 (Q), coordenada essa já utilizada nas seções anteriores.

Figura 9: Probabilidades de formação de contatos e Funil de estruturas. Funil de estruturas, construído utilizando as estruturas retiradas da dinâmica de enovelamento. Cada ponto representa uma configuração conformacional e as distâncias entre os pontos determinam a similaridade entre elas. Esse funil é colorido de acordo com a fração normalizada dos contatos nativos (QGRB2). Para complementar os resultados, é mostrado os mapas de contato, destacando os ensembles com cada mapa correspondente. Os mapas foram coloridos de acordo com a probabilidade de formação de contato (Pij). passando de poucos contatos formados e domínios desdobrados (azul) para contatos formados e domínios dobrados (vermelho).

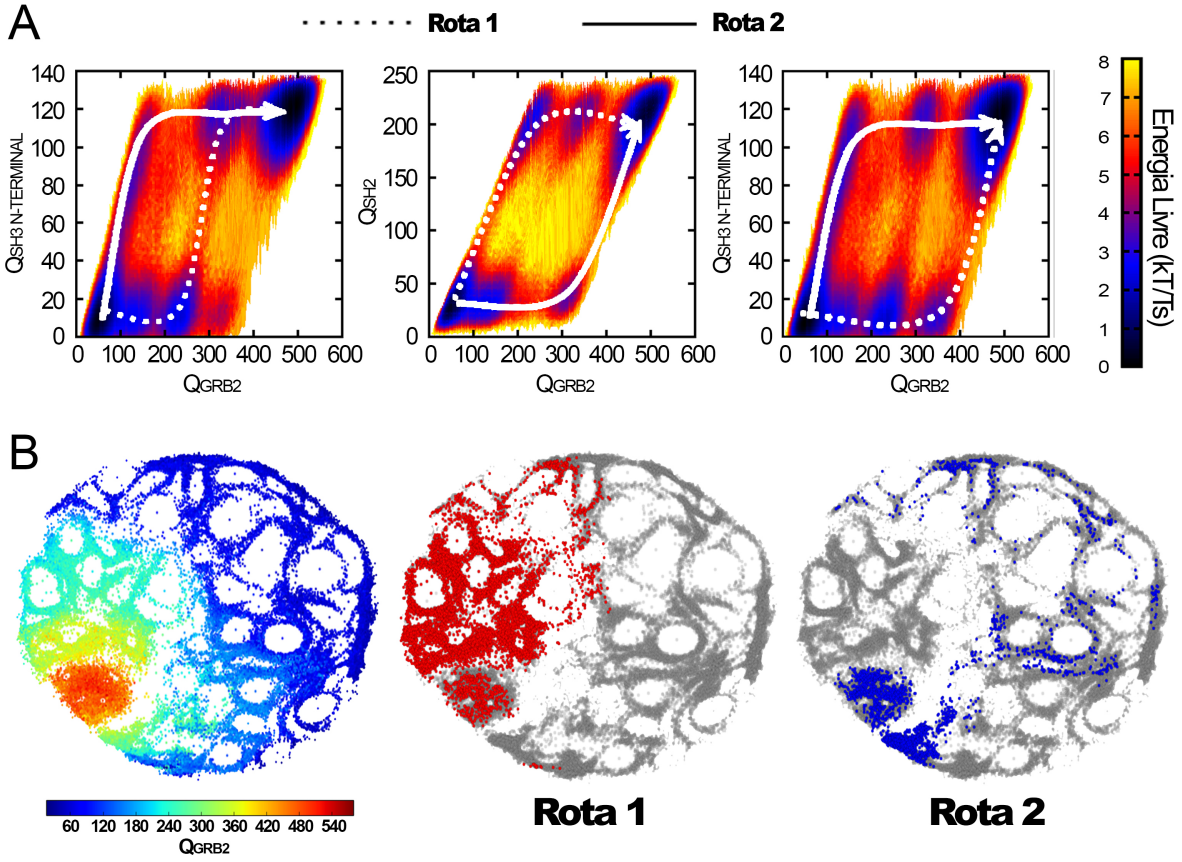


Fonte: Elaborada pelo autor (2022).

Combinando o funil de ensembles com o cálculo da probabilidade na formação dos contatos de cada ensemble, destaca-se aqui 6 regiões, cada uma correspondendo a uma combinação diferente de domínios estruturados: (1) Estados totalmente desenovelados; (2)

domínio SH3N parcialmente enovelado; (3) domínios SH3s enovelados; (4) domínio SH2 enovelado; (5) domínios SH2 e SH3N enovelados e (6) correspondendo às conformações nativas.

Figura 10: Caminhos de mínima energia. A) Superfície de energia livre construída utilizando os contatos de cada domínio de GRB2 como coordenadas de reação. Os valores de Energia Livre seguem a escala da barra de cores apresentada abaixo de 0 (preto) a 8 (amarelo) normalizados pela temperatura do estado de transição (Ts). B) Espaço de fase conformacional de GRB2 visualizado através do ELViM em função dos contatos nativos Q, e com as diferentes rotas destacadas em vermelho e azul.



Fonte: Elaborada pelo autor (2022).

Com essa nova visão em relação ao enovelamento de GRB2, a hipótese da ocorrência de backtracking na formação estrutural de seus domínios é descartada. Não há uma quebra de contatos para a formação de novos, e sim duas possibilidades nos caminhos de enovelamento, que se sobrepunham nas análises anteriores devido as coordenadas de reação Q e RMSD não conseguirem distinguir cada meta-estado de GRB2 (semi-enovelados). Para complementar e confirmar a existência das duas possibilidades nos caminhos de enovelamento, fizemos uma segunda análise usando frames consecutivos para mapear com precisão o caminho entre os

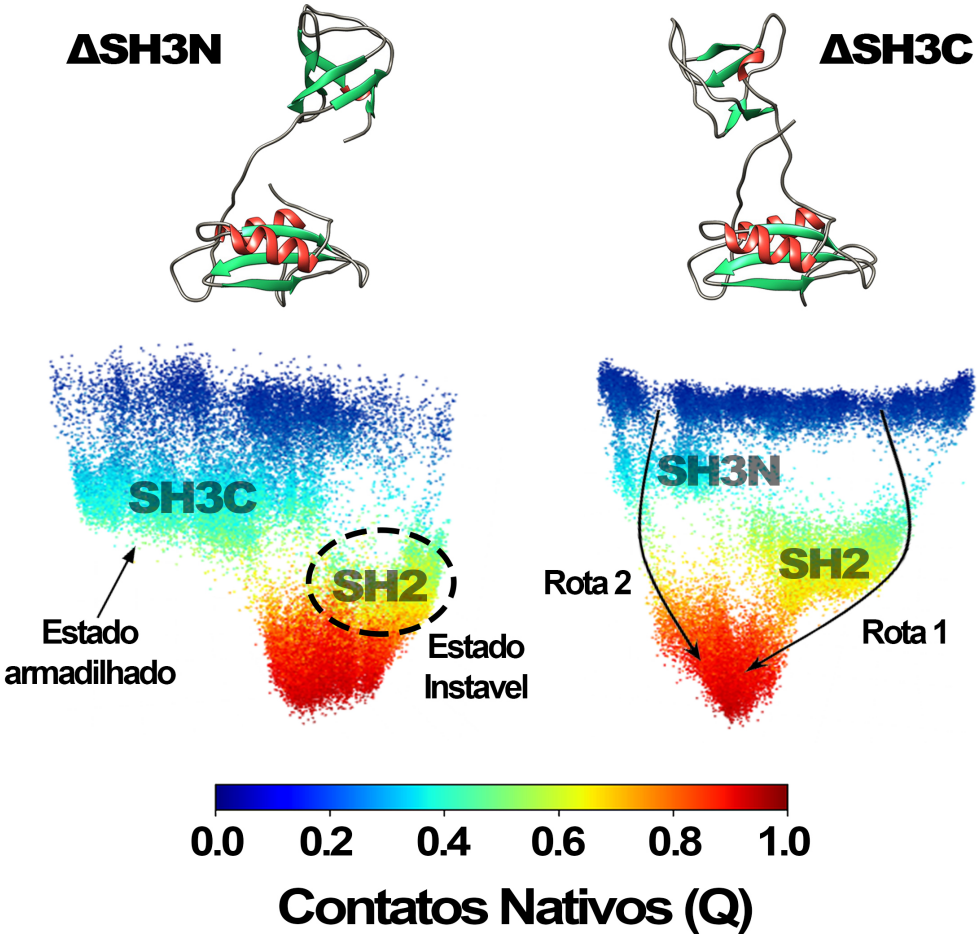
estados enovelados e desenovelados. Ainda, usamos o perfil de energia livre separando as coordenadas por domínios para calcular o caminho de mínima energia livre para cada uma das formações como mostrado na figura 10.

3.3.4 Deleção de domínios e estabilidade termodinâmica

Recentemente, Saki Ochi et. al. (63) mostrou experimentalmente que a deleção dos domínios interfere na estabilidade de GRB2. Resumidamente, a deleção do domínio SH3N diminui a estabilidade termodinâmica de GRB2, enquanto que a deleção de SH3C não tem esses efeitos. Utilizando nosso modelo e todos nossos resultados obtidos até então, decidimos construir duas novas estruturas, excluindo os domínios SH3N e SH3C e verificar se esses efeitos podem estar ligados com os caminhos de enovelamento da GRB2. Para isso, refizemos as dinâmicas moleculares com essas novas estruturas, calculamos o perfil termodinâmico e utilizamos mais uma vez o Software ELViM, como mostrado na figura 11.

Esses resultados indicam que a falta do domínio SH3N diminui a acessibilidade da rota 1 (SH2 – SH3C), ficando armadilhada no *ensemble* apenas com o domínio SH3C formado sem conseguir formar o domínio SH2. Quanto à deleção do domínio SH3C, os caminhos de transição não são afetados, a menos do domínio deletado. Esse conjunto de informações estabelece a importância da conexão interdomínios para a formação estrutural de GRB2. O conjunto de informações obtidas do enovelamento de GRB2 monomérica, toda a complexibilidade e a necessidade de todos os domínios estarem em contatos para a sua formação estrutural, despertou o interesse na avaliação de cada domínio separadamente, investigando a formação estrutural intra-domínios e possíveis interferências provenientes da comunicação inter-domínios.

Figura 11: Deleções de domínios e efeitos nos caminhos de enovelamento. Estruturas usadas para dinâmica molecular, coloridas de acordo com a estrutura secundária, com o verde representando as folhas$-\beta$ e laranja para as α-hélices. Os funis de energia mostram que a deleção do domínio SH3 $N_{TERMINAL}$ afeta os ensembles de estruturas, armadilhando a dinâmica no mínimo em que há apenas o $C_{TERMINAL}$ formado, desestabilizando a formação do domínio SH2. Já a deleção do SH3 $C_{TERMINAL}$ não afeta as rotas ou desestabiliza a formação de domínios.



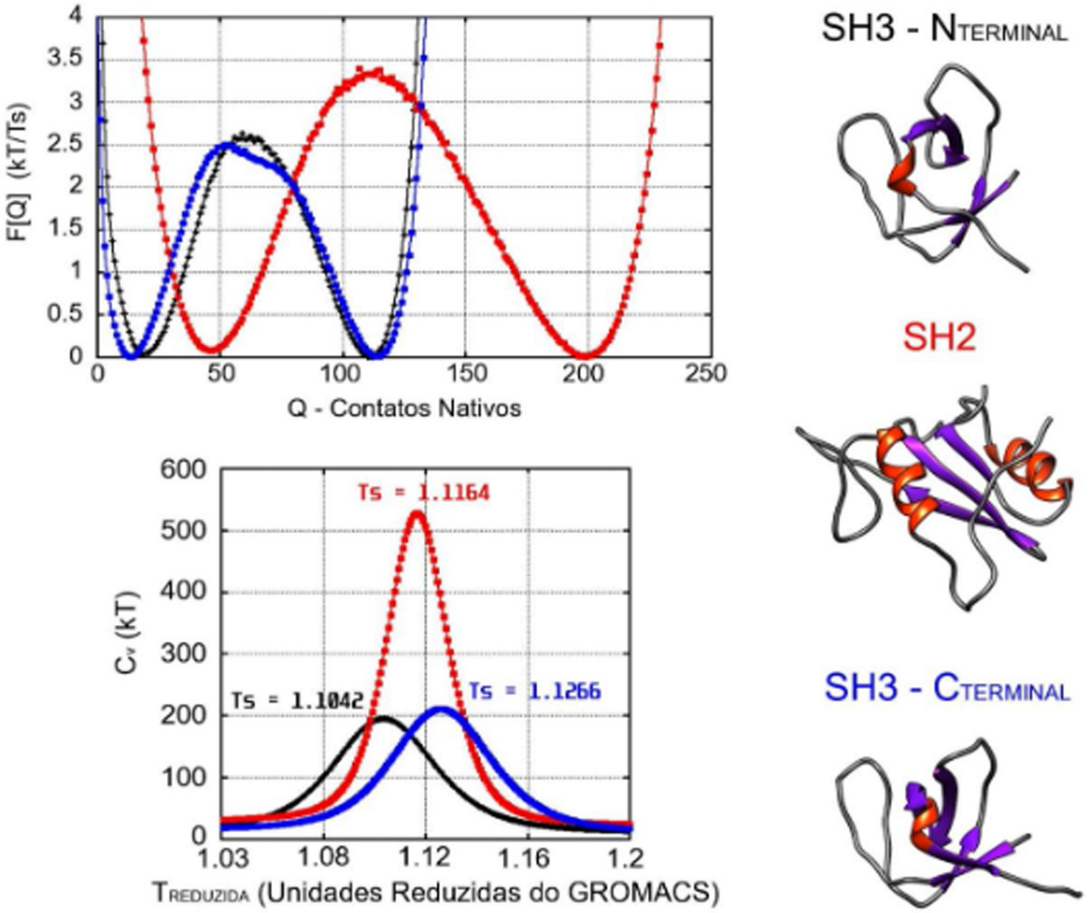
Fonte: Elaborada pelo autor (2022).

3.3.5 Perfil termodinâmico dos domínios de GRB2

Como já mostrado anteriormente, a proteína GRB2 é composta por três diferentes domínios: SH3N, SH2 e SH3C. Assim, separamos os domínios em diferentes arquivos de topologia e realizamos a dinâmica molecular de cada um desses. Na figura 12 é mostrado o perfil de energia livre de individual, sendo em preto referente ao domínio SH3N, vermelho SH2 e azul SH3C. O resultado mostra que o comportamento de cada um deles é de um enovelamento com dois estados e sem intermediários, a não ser pelo domínio SH3C, onde é verificado um pequeno ombro na barreira de energia, o que não chega a ser um estado

intermediário mas algum estado que está interferindo na energia livre do sistema.

Figura 12: Parâmetros termodinâmicos dos domínios de Grb2. Na parte inferior, está sendo mostrado a curva de calor específico pela temperatura em unidades reduzidas, os valores destacados para cada curva referem-se à temperatura do pico e também ao valor utilizado para cálculo da curva de energia livre mostrado na parte superior da imagem. A coordenação de reação utilizada são os contatos nativos (Q), o qual permite acompanhar as estruturas formadas ao longo do enovelamento. Também são mostrados cada domínio coloridos por suas estruturas secundárias: em roxo as folhas betas, laranja hélice alfas e cinza random coils, e seus nomes com cores referentes a cada curva nos gráficos de CV e Energia Livre.



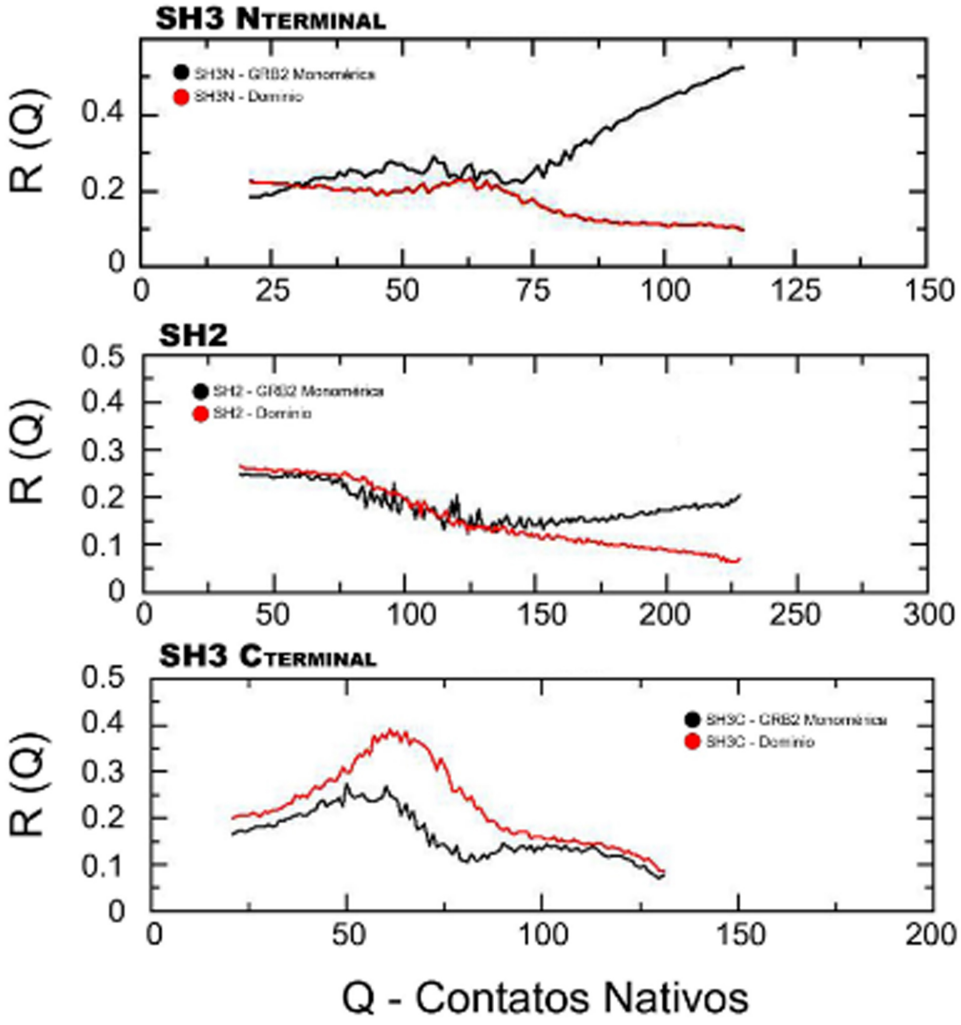
Fonte: Elaborada pelo autor (2022).

A primeira pergunta investigada acerca do assunto foi se a especificidade de formação de cada domínio era diferente quando tratado só como domínio isolado e quando inserido no monômero de GRB2. Para isso, calculamos a rota de enovelamento de ambos os casos como mostrado na figura 13. As curvas em vermelho referem-se aos domínios isolados e as curvas em preto aos domínios fazendo parte do monômero. Nos casos do domínio SH3N e SH2, as curvas são bem parecidas a não ser pelo valores finais da coordenada de reação de cada um, onde verifica-se uma subida na especificidade de ambos os domínios. Isso é

explicado pelo fato de que o domínio isolado tem um maior grau de liberdade nas suas extremidades. Quando inserido no monômero, sabemos que a extremidade C terminal de ambos os domínios são ligados covalentemente com os linkers que levam ao próximo domínio, diminuindo o grau de liberdade daquela região e interferindo nos possíveis movimentos para formação estrutural completa do domínio. Já para o domínio SH3C, quando tratado o domínio separadamente, notamos uma diferença no meio da curva, com aproximadamente 70 contatos formados (65% do total), mostrando uma possível diferença na formação de suas estruturas secundárias quando isolado.

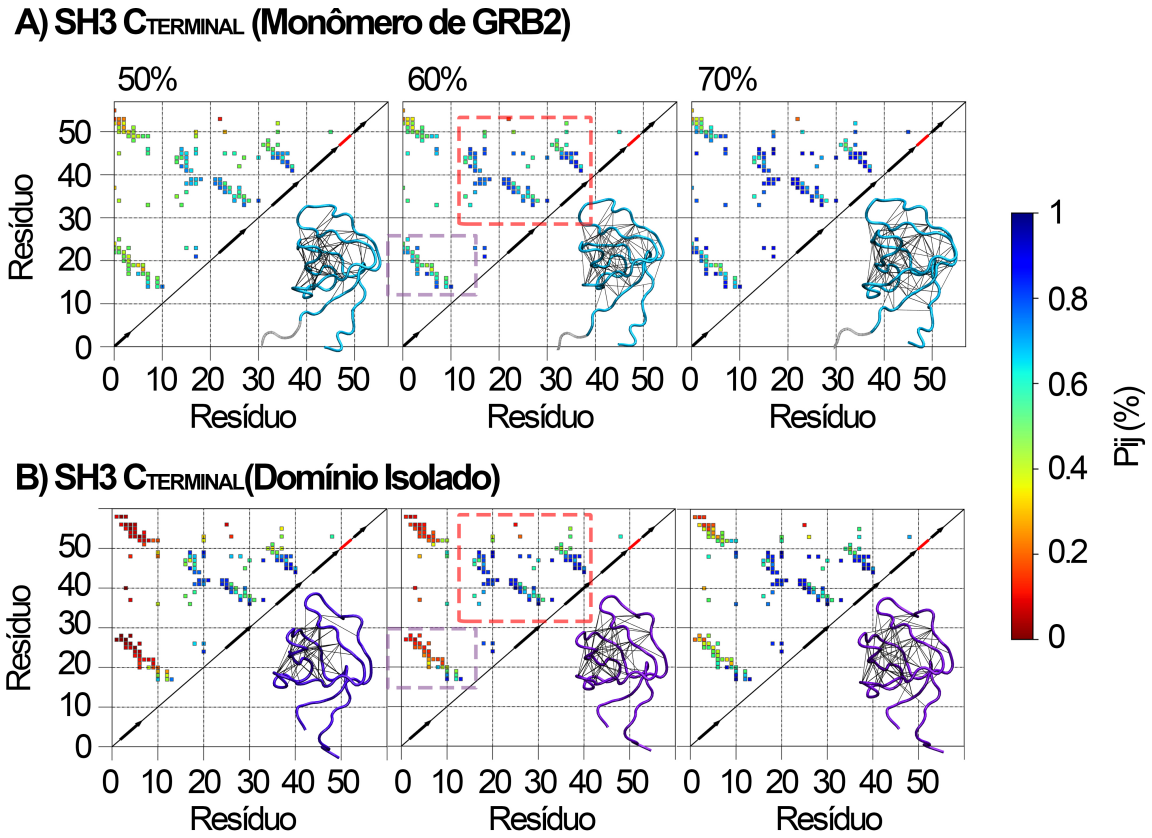
Para complementar esses resultados, foi calculado a probabilidade de formação dos contatos utilizando as dinâmicas de folding do domínio SH3C, Isolado e em monômero, com a pretensão de investigar quais as estruturas secundárias que estão sendo afetadas e consequentemente os resíduos envolvidos. Como mostrado na figura 14, as probabilidades são coloridas do vermelho para o azul (valores de 0 a 1) e estão sendo representados os valores de 60%, 70% e 80% dos contatos formados. O que notamos é que quando utilizamos o monômero a formação das betas 1 e 2 são simultâneos com a formação das outras estruturas secundárias do SH3C, já quando o domínio está isolado tem-se a formação das betas 3,4 e 5 e só depois de estruturadas que acontece a formação das betas 1 e 2.

Figura 13: Rotas de enovelamento dos domínios de GRB2. Estão sendo mostradas as rotas dos 3 diferentes domínios de GRB2 isolados e fazendo parte do monômero. Para os casos de SH3 $N_{TERMINAL}$ e SH2 o perfil no início das curvas diferem, tendo um valor mais alto para as curvas que representam os domínios ligados (monômero). Isso é explicado peloa fato de que as pontas C-terminais estão presas quando nessa forma, diminuindo o grau de liberdade das possibilidades e consequentemente aumentando a especificidade na formação das estruturas. Já para SH3 $C_{TERMINAL}$, há uma alteração nessa especificidade no intervalo médio dos contatos nativos.



Fonte: Elaborada pelo autor (2022).

Figura 14: Heterogeneidade na formação do domínio SH3 $C_{TERMINAL}$. A) Probabilidade na formação dos contatos nativos com 50,60 e 70% da estrutura formada utilizando as dinâmicas do monômero de GRB2 e B) utilizando as dinâmicas do domínio isolado. Nota-se aqui a formação conjunta das folhas betas do domínio SH3 $C_{TEMINAL}$ quando na forma monomérica de GRB2. Quando o domínio encontra-se isolado, essa formação não é heterogênea, sendo necessário a construção de algumas estruturas para depois finalizar a última beta. Os quadrados destacados no calculos de 60% podem ser utilizados para uma melhor visualização do fenômeno. As estruturas 3D nos cantos inferiores de cada gráfico ajudam a visualizar, sendo que cada linha entre os átomos representa as interações com probabilidade maior que 50%.



Fonte: Elaborada pelo autor (2022).

3.4 Conclusões

O ponto de partida para nossos estudos envolvendo GRB2 foi discutir sobre o enovelamento de sua estrutura na conformação monomérica, destacando-se os mecanismos envolvidos assim como outros parâmetros de interesse que possibilitam projetos futuros como a construção racional de fármacos.

Os primeiros resultados termodinâmicos demonstraram que nesse processo deveria haver estruturas intermediárias. Incrementando esses resultados com novas análises e parâmetros para a visualização dos fenômenos, mostrou-se que esses estados intermediários

são resultados da proteína ter multidominios com enovelamentos não concomitantes, formando-se estados semi-enovelados. Na continuidade dos cálculos, o funil de estruturas foi amostrado, nos dando uma perspectiva nova de um enovelamento com dois caminhos possíveis. No primeiro caminho mais provável, ocorre a formação do domínio SH2, seguido pelo domínio SH3 $N_{TERMINAL}$ e finalizando com a formação do domínio SH3 $C_{TERMINAL}$. Já a segunda possibilidade é a formação dos domínios SH3s primeiro e finaliza com a formação do domínio SH2.

Seguindo com as análises de enovelamento, também fizemos a deleção dos domínios para comparação direta com resultados experimentais da literatura, e o resultado obtido mostrou uma correlação entre o perfil de enovelamento com a estabilidade térmica de GRB2. Destaca-se aqui já a importância da comunicação inter-domínios para sua formação estrutural, sendo ponto de partida para o estudo dos domínios isolados. Com os domínios isolados, mostramos que o SH3 CTERMINAL tem seu perfil de enovelamento diferenciado quando trata-se do domínio e de quando está em conjunto com os outros fazendo parte do monômero de GRB2.

Todos esses resultados em conjunto mostram a complexibilidade de eventos que ocorrem durante o enovelamento de GRB2. Em conjunto com as funções de GRB2 destacadas na introdução deste capítulo, a importância desses estados semi-enovelados vai desde a flexibilidade estrutural que permite por exemplo mudanças dentro do dominios SH3s enquanto SH2 faz contato com alguma outra proteína envolvida em cascatas de sinalização, possibilitando o recrutamento de novas proteínas para o complexo. Além disso, essa estrutura flexível e com diferentes possibilidades permite inclusive a proteína agir em outras etapas dentro da célula, como por exemplo reparação de DNA, que já vêm sendo estudados e com resultados na literatura (64, 65)

A diferença estrutural na formação dos domínios separados e ligados (monômero) destaca a importância da comunicação entre os domínios para a formação da estrutura por completo. Uma vez que a estrutura é mais facilmente formada na presença dos outros domínios, podemos usar desses contatos para construção de moléculas que possam prevenir problemas de mau enovelamento ou até mesmo que facilitam o enovelamento e por consequência a sua estabilidade térmica.

Por fim, esses resultados nos trouxeram à hipótese de que possivelmente a GRB2 poderia agir como uma carreadora de pequenos fármacos com atividades já destacadas, como antiinflamatórios, anticarcinogênicas entre outras, para dentro de vias de sinalização.

Nessa perspectiva, nossos experimentos primeiramente foram de cálculos de interação como mostrado nos artigos em anexo deste trabalho.

CAPÍTULO II

Modelagem de Glicoproteínas

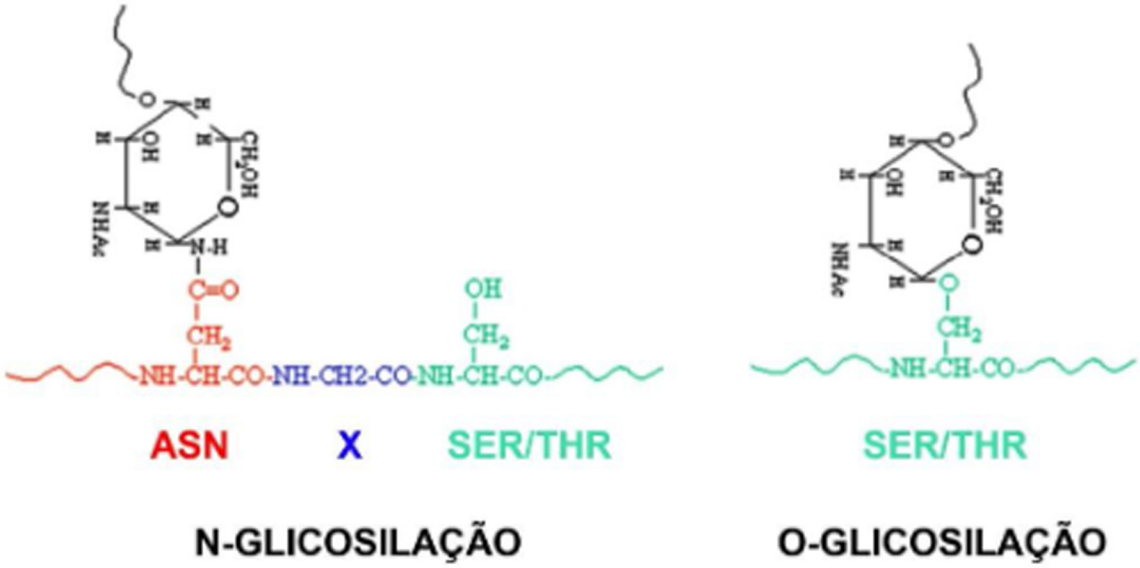
4 Glicoproteinas

4.1 Introdução às Glicoproteínas

As glicoproteínas fazem parte de um grupo de proteínas que são modificadas co-traducionalmente ou pós-traducionalmente com a adição de um ou mais oligossacarídeos a algum ou alguns aminoácidos específicos. Essas biomoléculas podem estar ligadas a uma grande variedade de processos biológicos, como por exemplo o crescimento celular, interação entre células, inflamação, defesa imunológica dentre outros.

O processo de associação dessas duas partes é denominado glicosilação. As glicoproteínas são classificadas de acordo com o tipo de ligação que ocorre entre as proteínas e os oligossacarídeos. Embora já descritos alguns tipos de interações (66, 67), as comumente encontradas em glicoproteínas são as N-glicosilações onde os oligossacarídeos são ligados no grupo amina de resíduos específicos de asparagina a resíduos de N-acetilglucosamina (GlcNAc) e as O-glicosilações através de grupos hidroxilas, principalmente em resíduos de serina ou treonina, via resíduos de N-acetilgalactosamina (GalNAc), manose (Man), xilose ou outros monossacarídeos (66–68), como ilustrado na figura 15.

Figura 15: Sítios de ligação de N-glicosilações e O-glicosilações. Na figura estão representados os dois tipos mais comuns de ligação dos glicanos em proteínas.



Fonte: Adaptado de *http://www.ionsource.com/Card/carbo/nolink.htm* (2022).

Vale ressaltar que, esses processos de glicosilação são altamente flexíveis, sendo possível

uma mesma proteína apresentar diferentes tipos de glicosilações, a depender unicamente de sua sequência e tamanho estrutural. Alguns estudos in silico indicaram que as glicosilações têm efeitos de estabilização dos resíduos ao redor do sítio de glicosilação e também em formação de estruturas secundárias (69). Também foram relatados efeitos de redução de área acessível ao solvente, quando a proteína está glicosilada (70). A Lectina pode ser usada como exemplo de proteína com efeitos positivos das glicosilações, com um aumento na sua estabilidade térmica em relação a sua forma não glicosilada (71–73).

A influência da glicosilação na manutenção da atividade biológica da proteína é bastante discutida, valendo citar, o papel dos glicanos na ajuda dos patógenos e células infectadas para escapar de reações imunes do hospedeiro (74). A preservação da atividade fosfodiesterase, estabilidade intracelular e proteção contra enzimas de degradação no caso da proteína SMPDL3A, que tem sua importância na regulação de macrófogos carregados de colesterol e com ações anti-inflamatórias. A glicosilação também impacta na ligação de receptores e manipulação das funções relatadas na porção Fc da Imunoglobulina (75), que inclusive é uma das proteínas de interesse nesse trabalho e que será melhor discutida na próxima seção.

Diversos testes de desnaturação e renaturação proteica vêm sendo realizados ao longo dos anos com a finalidade de relatar os efeitos dos glicanos para o enovelamento das proteínas, porém sem resultados conclusivos. Já é estabelecido que algumas dessas glicosilações têm efeitos intrínsecos no enovelamento, alterando seus perfis de energia livre, mas essa não é uma regra de validade geral. Os seus efeitos vão desde variação na estabilização de estruturas intermediárias até variações nas taxas de enovelamento. (75, 76).

Os efeitos relatados anteriormente resultam em sua maioria das modificações nas interações e estruturas devido ao glicano nas proteínas, entretanto, resultados obtidos com a Ribonuclease B revelaram que embora a N-glicosilação não altere a estrutura global ou local da proteína, o seu papel é apenas um redutor nas flutuações dinâmicas, promovendo um aumento na estabilidade. As estruturas globais das proteínas glicosiladas são quase idênticas às não glicosiladas, negligenciando o efeito da N-glicosilação na estrutura global da proteína (76).

Outras questões acerca das glicosilações vem sendo exploradas, como por exemplo a otimização racional enzimática via mutações de resíduos glicosilados, de modo que possam ser incluídos ou removidos na estrutura final a fim de alcançar as características desejadas para soluções biotecnológicas. Como exemplo, no estudo de uma proteína GH3 β-xilosidase,

onde é analisada uma gama de possibilidades para glico-mutantes, cada mutação com sua especificidade de aumentar ou não a atividade catalítica da proteína (77).

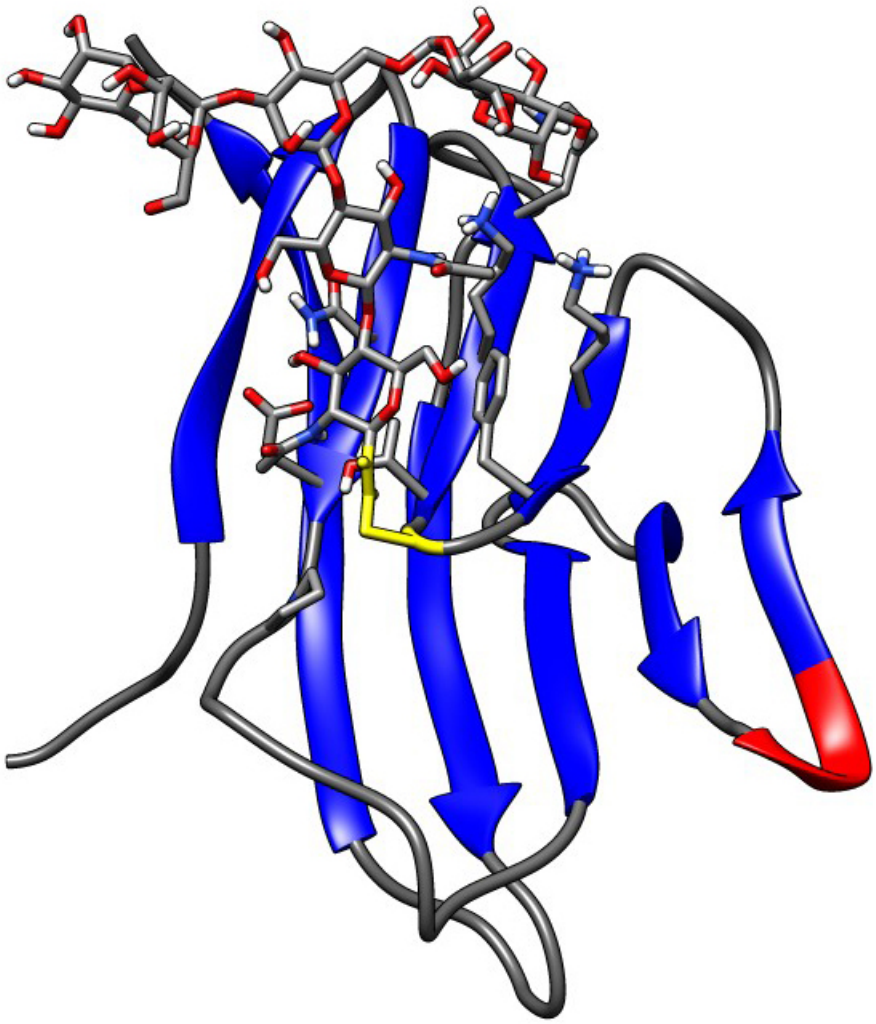
Ainda, recentemente o mundo presenciou (em 2019) os efeitos do SARS-CoV-2. As quatro proteínas estruturais do vírus SARS-CoV-2, spike (S), membrana (M), envelope (E) e nucleocapsídeo (N), estão sendo amplamente analisadas para determinar sua função na infecção por SARS-CoV-2 e gravidade da doença COVID-19. Dentre essas, 3 são glicoproteínas, onde os efeitos dos glicanos se mostraram diretamente correlacionados com a entrada viral nas células (78). Ainda estão sendo exploradas algumas abordagens para prevenir ou reduzir a biossíntese de N-glicano do hospedeiro e, assim, prevenir a glicosilação viral e possivelmente diminuir os efeitos do vírus (78).

4.2 Sistemas alvo de estudo

Visando explorar o efeito de glicosilações em diferentes situações, isto é, para diferentes proteínas e espécies, foi realizado um levantamento de sistemas de interesse. Buscamos em nossos achados uma generalização dos efeitos das glicosilações para diferentes casos de estudo, não os restringindo a um sistema específico, ainda que estejam sendo analisados deste modo. O primeiro sistema de interesse é o domínio de adesão da proteína CD2 humana (PDBid: 1GYA). CD2 humano é uma glicoproteína de superfície celular presente em linfócitos T e "*células natural killer*". Ela é importante para adesão celular e transdução de sinal por meio de interações com seu contra receptor CD58 (79, 80). Sua estrutura é composta por 105 resíduos de aminoácidos e contém apenas um único sítio de N-glicosilação em Asn-65 como mostrado na Figura 16.

A estrutura de CD2 foi determinada por RMN para o domínio glicosilado (80) e também para os componentes peptídicos e glicanos separados (79, 80). Esses resultados mostraram que não há diferenças significativas na conformação proteica com e sem glicano, e nem na composição dos glicanos. Logo, a presença do glicano não altera a conformação nativa de CD2. Todavia, estudos termodinâmicos forneceram uma análise energética quantitativa sobre o efeito da N-glicosilação na termoestabilidade e na cinética de dobramento de CD2. Os resultados experimentais mostraram que o sistema glicosilado é 3,1 kcal/mol mais estável e desdobra-se 50 vezes mais lentamente do que o seu homólogo não glicosilado (79). Outras análises também mostraram que CD2 deglicosilado, utilizando enzimas para o corte, bem como uma molécula de CD2 mutante contendo uma única substituição Asn65-Gln65

Figura 16: Estrutura da Proteína CD2. O domínio de adesão da proteína CD2 humana é representado juntamente com sua glicosilação. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de ASN65 em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho.



Fonte: Elaborada pelo autor (2022).

demonstra que CD2 deglicosilado não se liga à CD58 ou anticorpos monoclonais dirigidos a epítopos de domínio de adesão CD2 nativo, mostrando que a N-glicosilação é essencial para as funções de adesão ao CD2 humano (79).

Um segundo sistema de interesse é uma subunidade do homodímero IgG-FC (PDB id: 1H3W), também conhecida como Imunoglobulina. A estrutura da proteína Fc homodimérica

tem uma forma chamada de “ferradura” que consiste em duas subunidades contendo cada uma delas um domínio CH2 e CH3. Essas subunidades interagem via uma ligação dissulfeto nos domínios CH3/CH3. Seu sítio de N-glicosilação N297 é encontrado nos domínios CH2, como mostrado na Figura 17 (81, 82). Uma das características física e estrutural relacionadas com a posição desse glicano é seu preenchimento da lacuna entre os domínios CH2, “protegendo” as superfícies hidrofóbicas localizadas internamente ao homodímero do solvente (82). Na literatura é amplamente discutido o papel dos açúcares para a conformação estrutural do homodímero. A proteína teve sua estrutura calculada com e sem glicosilações e conformações mais abertas ou fechadas são moduladas pela presença e tamanho dos glicanos (82). Essa modulação é importante para reconhecimento e interação com outras macromoléculas, ditando inclusive a efetividade em suas atividades. Em relação ao seu enovelamento e estruturas da proteína, experimentos mostram uma desnaturação de dois estados devido a presença de dois diferentes domínios, i.e, enovelamento não concomitante de seus domínios. Todavia, a glicosilação não tem efeitos, sendo o mesmo enovelamento para a proteína com e sem glicanos (83).

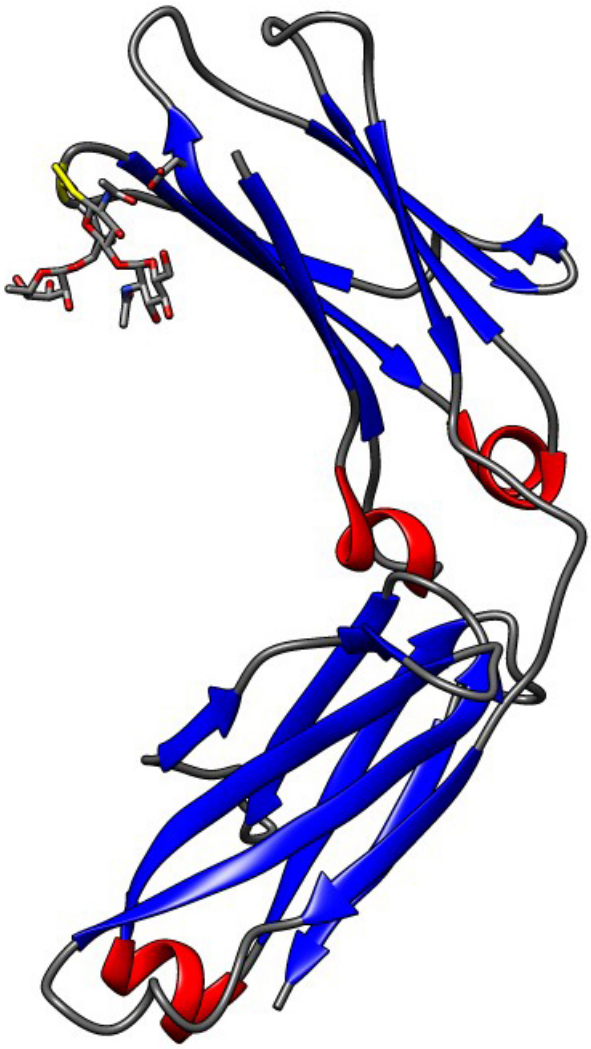
A terceira glicoproteína é conhecida como Erns do pestivirus BVDV-1, sua estrutura proteica é um monômero com 6 sítios N-glicosilados, como mostrado na Figura 18. Seu C-terminal é relatado com múltiplas funcionalidades como: âncora para membrana, sinalizador de retenção/secreção, sítio de ligação para glicosaminoglicanos (GAGs) da superfície celular e um local de clivagem de peptidase de sinal (84).

Cada um dos sítios de N-glicosilação foi usado para investigar seus efeitos na virulência em suínos (85). Os resultados desses estudos in silico mostram que apenas uma das remoções da N-glicosilação produziu vírus atenuados, que ao contrário dos outros mutantes, levando a uma infecção transitória em suínos caracterizada por sintomas leves e diminuição da disseminação do vírus, sugerindo que a glicosilação poderia ser modificada para o desenvolvimento de vacinas (85).

Outros estudos da proteína Erns, mostraram que a deglicosilação in vitro não resultou em uma perda importante de sua atividade (85, 86). No entanto, uma segunda análise foi necessária uma vez que a deglicosilação foi realizada em uma proteína secretada totalmente enovelada. Nessa segunda análise, é utilizada uma abordagem in vivo inibindo as a-glicosidases I e II residentes no retículo endoplasmático, mostrando que os glicanos não são necessários para o enovelamento da proteína, contudo necessário para uma secreção eficiente das proteínas (85, 86). Este efeito não é causado por mudanças de conformação, uma vez

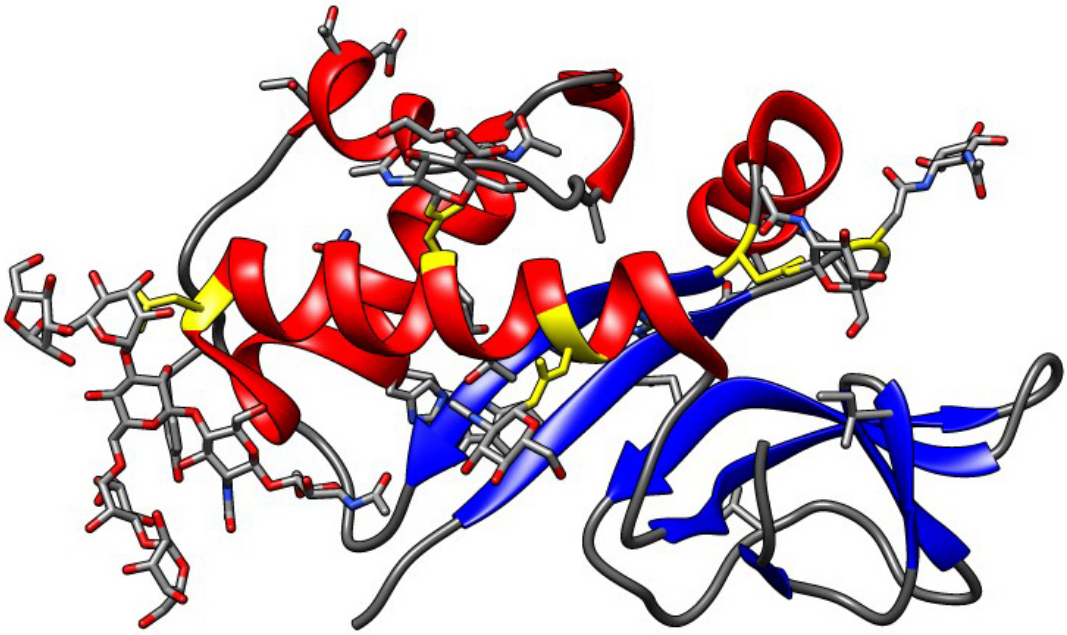
que cessou/suprimiu totalmente sua atividade enzimática na presença do inibidor, mas sim devido a um envolvimento direto dos N-glicanos ligados no transporte de a proteína pela via secretora (86).

Figura 17: Estrutura da Proteína IgG-FC. Também conhecida como Imunoglobulina, IgG-FC é representado juntamente com sua glicosilação. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de Asn em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho.



Fonte: Elaborada pelo autor (2022).

Figura 18: Estrutura da Proteína Erns. A proteína viral Erns é representada juntamente com suas glicosilações. Seus domínios foram coloridos de acordo com sua estrutura secundária (azul - folhas betas, vermelho - hélices alfas e cinza - random coils) e o resíduo de Asn em que o glicano é ancorado também está destacado na cor amarela. Os carbonos do glicano são coloridos em cinzas com os hidrogênios destacados em vermelho.

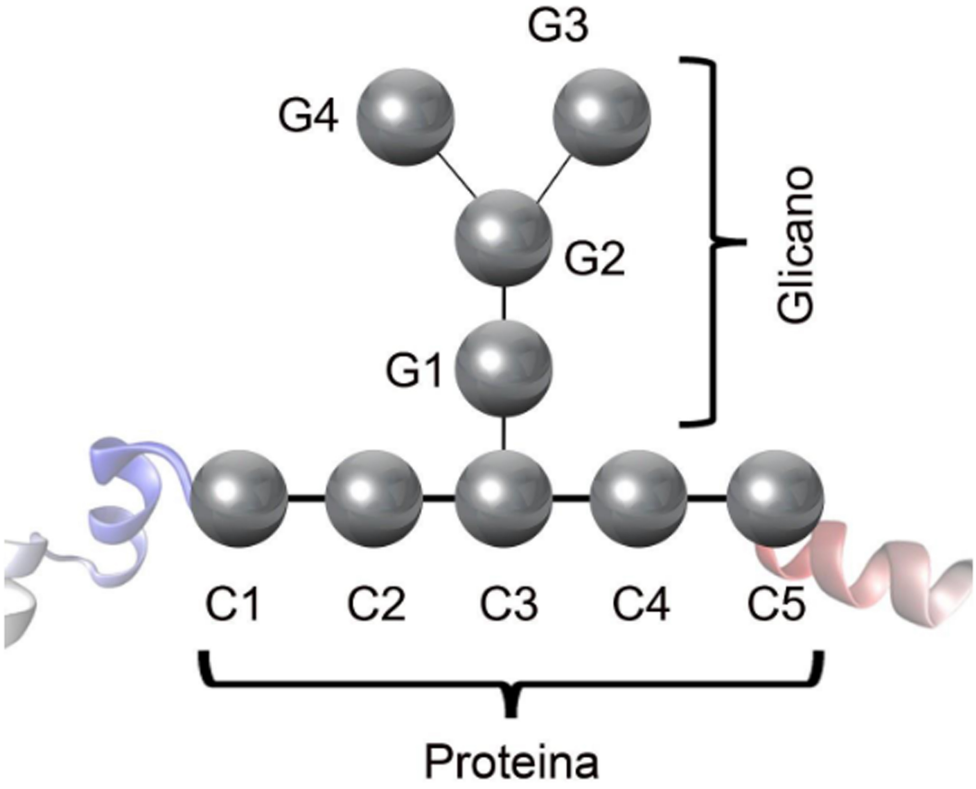


Fonte: Elaborada pelo autor (2022).

4.3 Modelagem do sistema

A abordagem utilizada para modelagem das glicosilações utilizando o modelo baseado em estrutura Cα, é representar os anéis dos sacarídeos por apenas um Carbono central, como mostrado na figura 19.

Figura 19: Modelagem da glicosilação com o SBM. Um esquema de glicosilação simples é mostrado, com apenas 4 anéis sendo representados. As esferas nomeadas com C representam os carbonos da proteína em que está ligado o açúcar, as nomeadas com G representam o próprio açúcar e os índices indicam os resíduos ou anéis de sacarídeos.



Fonte: Elaborada pelo autor (2022).

Os termos de potenciais alocados para os resíduos da glicosilação foram os de ligações covalentes, ângulos, diedros e repulsivo (restrição estérica). Os potenciais de interações (contatos) não foram adicionados para essas glicosilações, fazendo com que os glicanos tenham seus movimentos livres, de modo que a única interação com as outras moléculas seja apenas repulsiva, impedindo a sobreposição de átomos. Esse tipo de modelagem é o mesmo adotado na referência (77), dentre outros grupos de pesquisa. Um esquema de como são essas adições encontra-se na Figura 20.

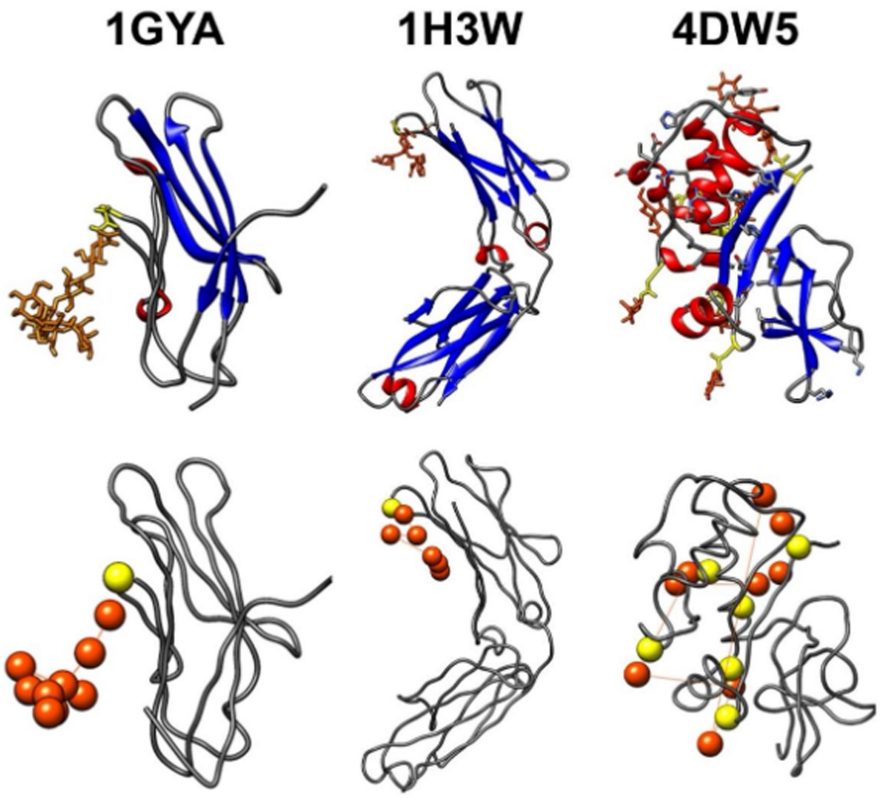
Figura 20: Adição dos parâmetros na topologia da proteína. Para um esquema como o que foi mostrado na figura 19, as novas combinações de resíduos com a adição do glicano é representada. Essas combinações representam os novos parâmetros adicionados no arquivo de topologia e que representam a presença do glicano na estrutura proteica.

Covalentes:	
C3 - G1	G1 - G2
G2 - G3	G2 - G4
Ângulos:	
C2-C3-G1	C4-C3-G1
C3-G1-G2	G1-G2-G3
G1-G2-G4	
Diedros:	
C1-C2-C3-G1	C5-C4-C3-G1
C2-C3-G1-G2	C4-C3-G1-G2
C3-G1-G2-G3	C3-G1-G2-G3

Fonte: Elaborada pelo autor (2022).

Para as três proteínas utilizadas no trabalho, todas as modelagens foram feitas utilizando suas estruturas resolvidas, incluindo o glicano, em seus tamanhos e posições (relativo ao centro de massa), como mostrado na figura 21.

Figura 21: Modelagem das proteínas glicosiladas. Na parte superior estão representadas as proteínas coloridas por suas estruturas secundárias (hélice alfa em laranja, folhas betas em azul e random coil em cinza) juntamente com as glicosilações em laranja. Na parte inferior, as proteínas já modeladas por seus carbonos alfa e glicosilações, em amarelo estão representados as ASN em que estão ligados as glicosilações em laranja.



Fonte: Elaborada pelo autor (2022).

4.4 Objetivos

Avaliar o papel das glicosilações no enovelamento de três glicoproteínas distintas com estruturas cristalográficas resolvidas utilizando nossa modelagem simplificada. A escolha de três glicoproteínas se justifica pelos diferentes relatos de efeitos encontrados na literatura. Com a avaliação com os efeitos individuais poder encontrar características globais relacionadas às glicosilações, assim como entender o seu papel evolutivo.

4.5 Resultados e discussões

4.5.1 Perfil de Energia Livre dos sistemas

Para avaliar a influência das glicosilações no processo de enovelamento proteico utilizamos o cálculo do calor específico a volume constante e dos seus perfis de Energia Livre para cada uma das proteínas avaliadas. Para isso, foram utilizadas as coordenadas de reação: número de contatos nativos formados (Q) e RMSD.

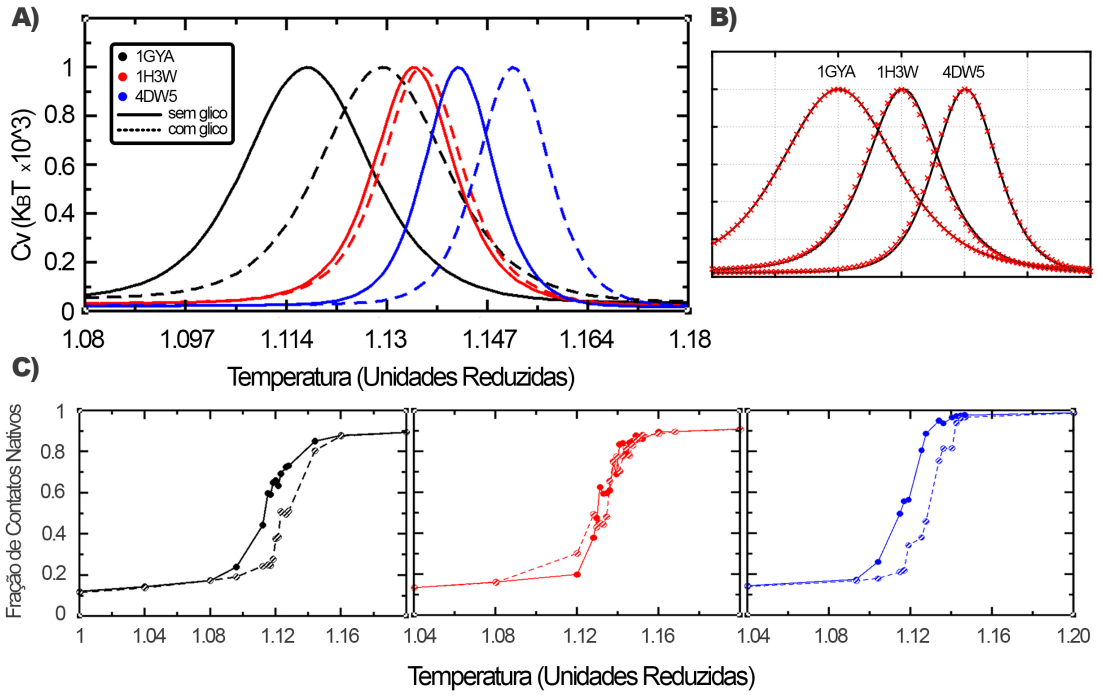
A figura 22 mostra que para todos os casos, a curva do valor específico foi deslocada para a direita, mais evidencialmente para as proteínas CD2 (1GYA) e Erns (4DW5). Esse deslocamento pode indicar um possível aumento em suas estabilidades térmicas. A sobreposição dessas curvas mostrada na figura 22B mostra que não há alargamento das distribuições em nenhum dos três casos, descartando hipóteses sobre influência na cooperatividade pela presença do glicano.

A fração de contatos nativos é uma média sobre todos os frames das simulações para cada uma das temperaturas utilizadas. Valores próximos de 1, significa uma dinâmica majoritariamente com a proteína enovelada, valores próximos de 0 para casos desenovelados e intermediários para dinâmicas nas temperaturas de transição, como mostrado na figura 22C. Mais uma vez, para as proteínas CD2 e Erns há um deslocamento, mostrando que existe a influência do glicano para estabilidade térmica de ambas.

Os cálculos de Energia livre mostram que para a proteína CD2, o resultado sugere uma transição de dois estados, com a glicosilação afetando diretamente na barreira de energia (vista apenas pela coordenada RMSD). Para a Imunoglobulina, ocorre uma transição com um estado semi-enovelado, devido à presença dos dois diferentes domínios e não serem enovelados concomitantemente, suas barreiras de energias tanto para o caso glicosilado quanto para deglicosilado são iguais. Já para a proteína Erns, a transição passa por um estado intermediário, e que é mais pronunciado quando temos a glicosilação como mostrado na Figura 23. Em termos energéticos, também há uma diminuição na energia necessária para acessar o estado intermediário e um aumento na barreira de transição para os estados desenovelado.

Esses resultados envolvendo o perfil de energia livre e curvas de calor específico estão de acordo com os resultados experimentais da literatura citados na introdução deste capítulo, complementando-os. Foi possível notar inclusive os mesmos efeitos sob o ponto de vista

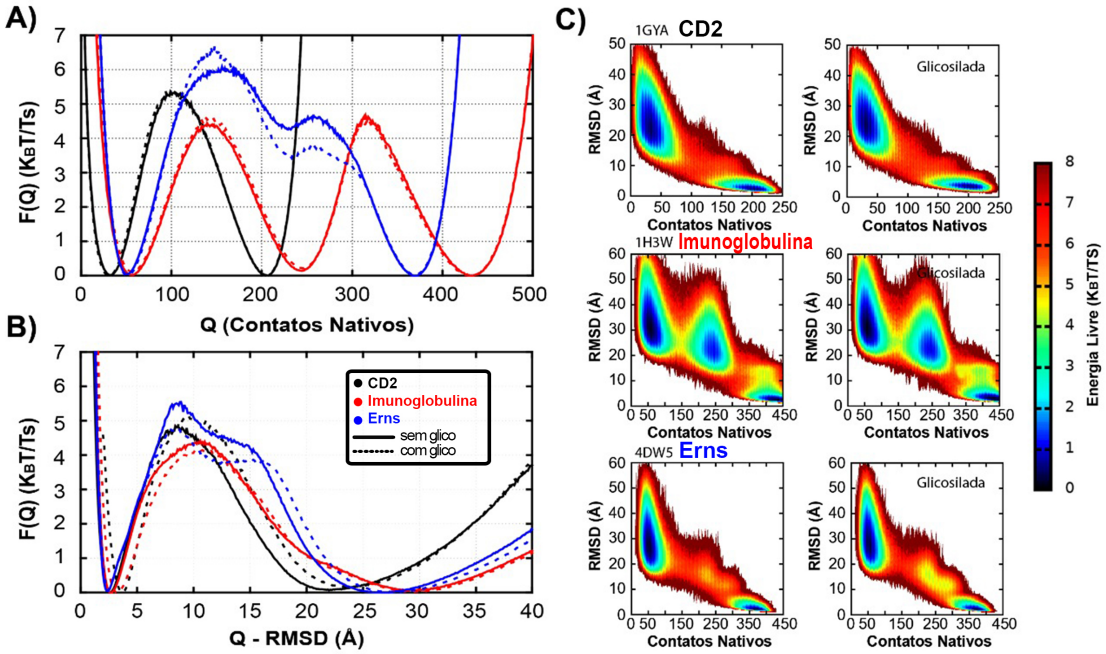
Figura 22: Calor Específico a volume constante. (A) Calor específico x Temperatura em unidades reduzidas das proteínas estudadas. (B) Sobreposição das curvas contínuas e tracejadas mostrada em A. (C) Fração de contatos nativos para cada uma das proteínas estudadas. Para todos os 3 casos, as curvas contínuas e tracejadas correspondem às estruturas sem e com glicosilações respectivamente. Com a adição das glicosilações, as proteínas PDB ids: 1GYA e 4DW5 tiveram suas curvas de Cv e Fração de Contatos desviadas para a direita, evidenciando um possível aumento em sua estabilidade térmica. As sobreposições mostram que não há um alargamento das curvas, descartando hipóteses de cooperatividade provindas da adição de glicosilações.



Fonte: Elaborada pelo autor (2022).

computacional para esses sistemas biológicos, ressaltando a consistência do modelo. Deste modo o papel biológico é de extrema importância, e deve ser considerado para justificar as múltiplas indicações da função da glicosilação.

Figura 23: Energia Livre para os casos glicosilados e deglicosilados. (A, B) Perfil de energia livre para a coordenada de contatos nativos e RMSD respectivamente. (C) Superfície de energia utilizando a correlação de ambas as coordenadas. Para todos os casos, as curvas contínuas e tracejadas correspondem às estruturas sem e com glicosilações respectivamente. As curvas em preto, representando a proteína CD2, mostra que a adição da glicosilação afeta a barreira de energia livre (vista apenas para a coordenada RMSD) e é uma transição de dois estados. Para a proteína Imunoglobulina (curvas em vermelho), não há alterações nas barreiras de energia e é uma transição passando por um estado semi-enovelado devido a presença dos dois diferentes domínios. Por fim, as curvas em azul representam a proteína Erns em que há um estado intermediário facilitado pela adição de glicosilação e um aumento na barreira para seu desenovelamento.



Fonte: Elaborada pelo autor (2022).

Detalhando, para as proteínas CD2 e ERNS viral onde foram relatados aumentos em suas termoestabilidade, nossos resultados conseguem por meio do calor específico encontrar diferenças na temperatura de transição. Já para o caso da Imunoglobulina, onde experimentalmente é calculado um enovelamento de dois estados e sem alterações na termoestabilidade influenciadas pelo glicano, nosso modelo além de conseguir visualizar essa transição de dois estados, também não mostrou diferenças significativas tanto no calor específico quanto no perfil de energia livre.

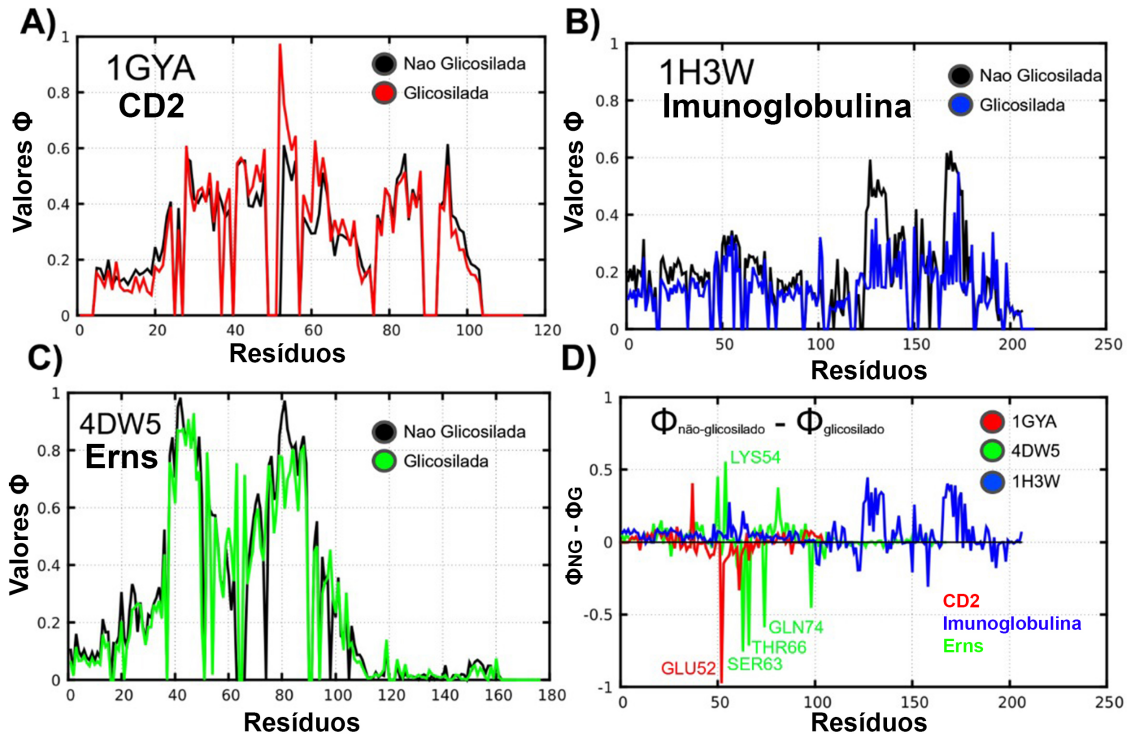
4.5.2 Avaliação das influências nos mecanismos de enovelamento por meio dos valores-ϕ

Uma das hipóteses para os efeitos da adição da glicosilação das proteínas, seria a alteração na relevância de alguns contatos, podendo influenciar diretamente na estabilidade dessas proteínas. Em nosso caso, a utilização dessa metodologia pode nos refinar alterações nos mecanismos de enovelamento pela influência das glicosilações. O resultado mostrado na figura 24 mais uma vez mostra que a glicosilação tem efeitos diferentes dependendo do sistema e de sua função biológica associada.

Para a proteína CD2, a adição da glicosilação aumenta a relevância dos contatos com os resíduos Glu52, tendo este uma variação de 0.98 em relação ao seu valor quando a proteína está deglicosilada (0). Para a IgG1, o cálculo mostra que os valores phi para a proteína sem glicosilação são maiores, e o fator mais intrigante, é que esses resíduos pertencem ao domínio não glicosilado. Apesar de relatado, esses valores não tem mudanças maiores que 0.4 e também nenhum desses resíduos tem seus valores phi próximos a 1 com ou sem glicosilações.

Para a proteína Viral Erns o efeito foi mesclado, sem a glicosilação os cálculos mostram que alguns resíduos, como por exemplo, Asp50, Lys54 e His81 eram importantes para o enovelamento da proteína. A adição da glicosilação teve como efeito diminuir esses valores, principalmente para o resíduo de Lys54 para o qual o valor diminui em aproximadamente 0.5. Os resíduos Ser63, Thr66 e Gln74 têm efeitos inversos, aumentando a importância desses em pelo menos 50% quando adicionado a glicosilação, sugerindo alterações no mecanismo de enovelamento.

Figura 24: Valores ϕ. (A,B e C) Curvas das proteínas com código PDB 1GYA, 1H3W e 4DW5 respectivamente. Nessas, as curvas com os valores phi com e sem glicosilação são sobrepostas, mostrando a influência da glicosilação em alguns contatos. As relações entre aumento e diminuição nos valores phi são mais evidenciadas nos gráficos (A) e (C), destacando-se alguns possíveis resíduos de importância para o estado de transição, como mostrado em (D).



Fonte: Elaborada pelo autor (2022).

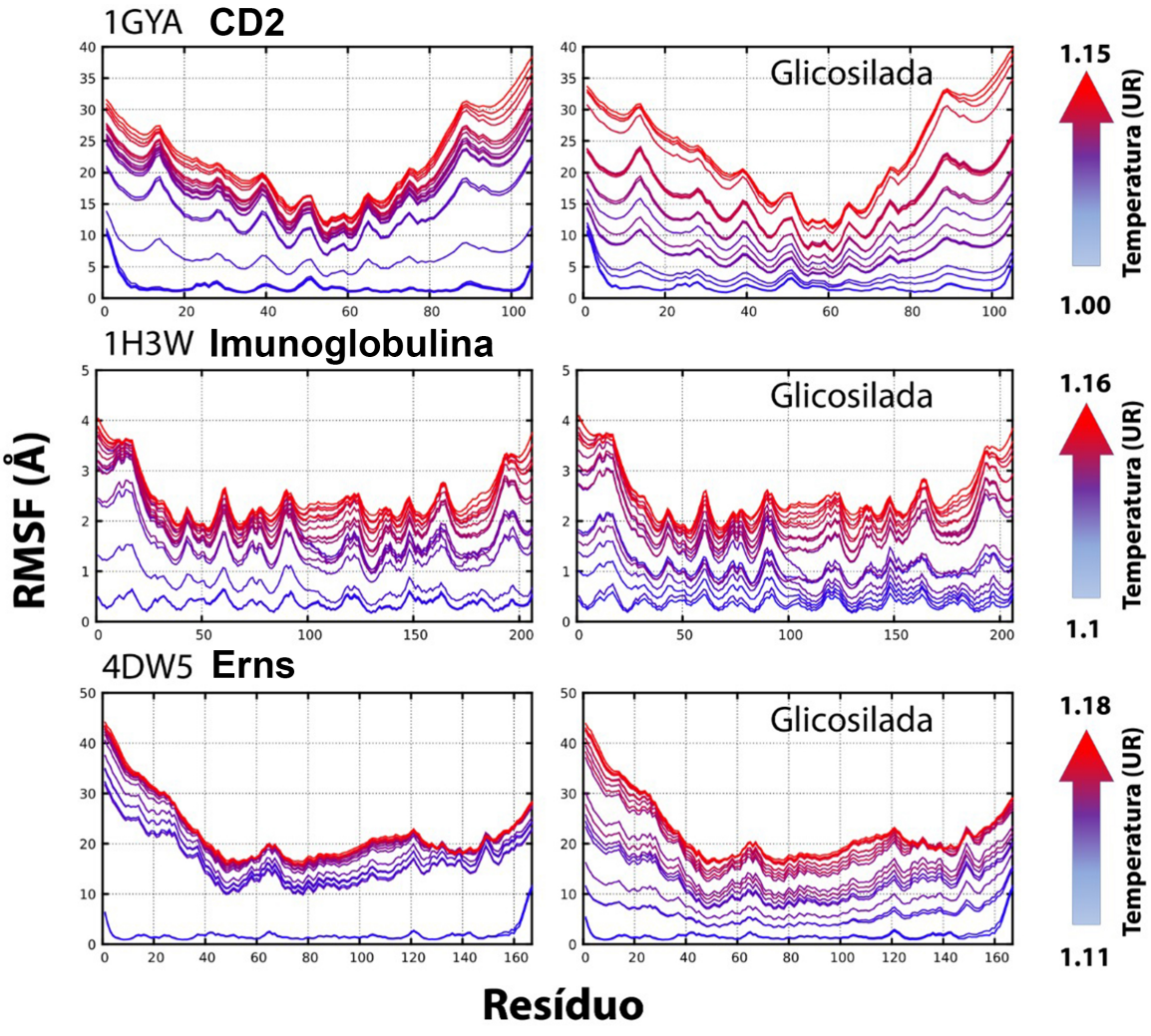
4.6 Flutuação média dos resíduos

Motivado pelos resultados anteriores, onde cada glicano tem seus efeitos próprios para cada caso, calculamos o RMSF de cada resíduo em cada temperatura simulada. Para esses gráficos foram utilizadas as temperaturas pré estado de transição, transição e pós transição com base nas curvas de calor específico/energia livre. O resultado mostrado na figura 25, mostra o regime de flutuação dos resíduos com e sem as glicosilações.

Nossa hipótese era que, com o aumento da temperatura no sistema essas flutuações aumentariam gradativamente, todavia foi possível notar variações no perfil para os diferentes sistemas. Vale ainda ressaltar que as maiores diferenças encontradas quando comparados os gráficos com e sem glicosilação por meio dos RMSFs, são para as proteínas CD2 e ERNS, as quais tanto experimentalmente quando por nossos resultados computacionais de calor específico tem uma maior estabilidade térmica quando adicionado os glicanos. Apesar da

proteína IgG2 também ter o comportamento diferente com e sem glicanos nas flutuações, essas diferenças não se mostram significativas. Esses comportamentos nos sugerem que a presença do glicano não interfere apenas na região vizinha ao local de interação. É uma interação expandida para toda a proteína de maneira que todos os resíduos sentiram os efeitos da presença do açúcar.

Figura 25: Perfil de flutuação (RMSF). Na esquerda são apresentados os perfis das proteínas deglicosiladas e na direita as proteínas glicosiladas. As cores são representadas pelas temperaturas das quais foram retirados o perfil de flutuação. Nota-se que a glicosilação teve como efeito preencher os estados com flutuações intermediárias.

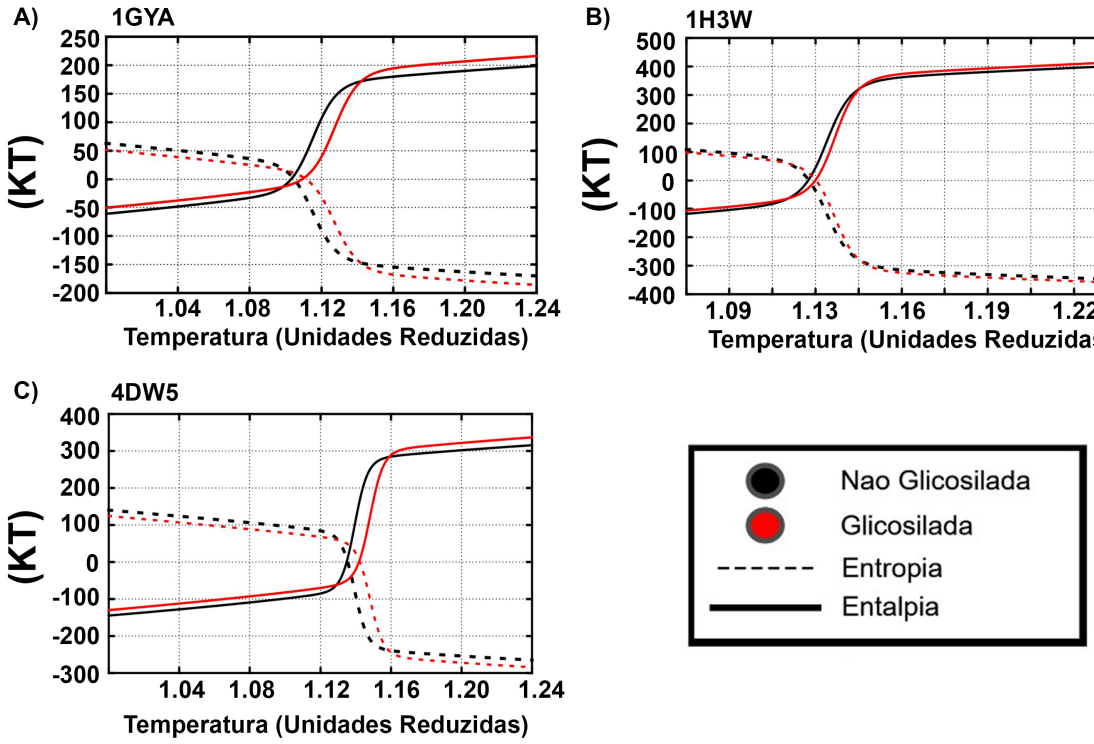


Fonte: Elaborada pelo autor (2022).

4.7 Entalpia e Entropia das transições

Com o objetivo de avaliar como a presença das glicosilações restringem o acesso às conformações (ou influenciam no direcionamento ao estado nativo), foram calculados por meio do WHAM as entropias e entalpias em função da temperatura para os três sistemas, com e sem glicosilação, como mostrado na figura 26.

Figura 26: Entropia e Entalpia do sistema. Os valores de entalpia e entropia estão sendo mostrados para as três proteínas de interesse. (A,B e C) separam os resultados das proteínas 1GYA, 1H3W e 4DW5 respectivamente. Cada gráfico contém as curvas de Entalpia e entropia, diferenciando-se pela linha tracejada e contínua. E também para cada gráfico é apresentado seus valores tanto na sua forma deglicosilada quanto glicosilada, representados pelas cores preta e vermelha.



Fonte: Elaborada pelo autor (2022).

A adição do glicano na proteína altera os valores energéticos de entropia e entalpia para os casos de 1GYA e 4DW5, entretanto o balanço entre entropia configuracional e entalpia do sistema não tem alterações. Essa mudança energética é correlacionada às suas estabilidades térmicas e que por sua vez tem origem na desestabilização dos estados desenovelados. Nosso modelo e os resultados obtidos sugerem que a presença do glicano força o conjunto desenovelado a adotar um número maior de configurações estendidas quando comparadas ao sistema sem glicosilação.

4.8 Conclusões

A diversidade de efeitos que vem sendo relatado sobre a influência das glicosilações nos estudos de proteínas traz consigo uma infinidade de questionamentos biológicos, bioquímicos e biofísicos. Aqui neste trabalho, foi explorado por modelos simplificados a influência da glicosilação no processo de enovelamento para três proteínas distintas.

Os resultados termodinâmicos, com perfis de Energia Livre e calor específico mostram que nosso sistema modelado tem resultados diferentes devido a adição da glicosilação em cada proteína. Duas das proteínas estudadas, CD2 e ERNS têm resultados in silico de aumento em suas termoestabilidade, sendo essas mesmas duas proteínas as quais mostraram diferenças em nossos resultados teóricos de Calor Específico. A proteína em que não se obteve diferenças teóricas em relação às temperaturas de transição, também não tem resultados in vivo a respeito de melhoramento térmico.

Já a respeito das energias livres, o único perfil com alteração significativa em nossos resultados foi o da proteína viral ERNS, em que um estado de transição foi mais populado com a adição da glicosilação. Ainda que não haja correlação sólida, essa é a única proteína estudada em que foi mostrado efeitos na secreção in vivo.

Uma segunda análise, utilizando valores ϕ, também seguiu a linha de que cada sistema reage de forma diferente com adição da glicosilação. Entretanto, uma conclusão geral de acordo com dois de nossos três sistemas estudados, as variações mais significativas nos valores phi ocorrem com a adição da glicosilação, i.e, a glicosilação influência aumentando a importância de alguns resíduos para o enovelamento.

Os resultados encontrados indicam que com uma modelagem simples, como é o Cα SBM é possível prospectar os efeitos da glicosilação em uma proteína, servindo assim como ferramenta para o desenho racional de características desejáveis por meio da inserção de glicosilações.

Os valores de RMSF para cada temperatura simulada juntamente com os valores de entropia e entalpia ao longo das simulações mostram que um dos efeitos da glicosilação é afetar o balanço energético a fim permitir maior liberdade conformacional ao longo do enovelamento, em acordo com os cálculos de valores de phi. Esse é um processo visto em nossos três sistemas, ainda que um pouco menos significativo para a IgG1, mas que está diretamente relacionado com os resultados anteriores.

Referências

1 VOET, D.; VOET, J. G.; PRATT, C. W. *Fundamentos de Bioquímica-: A Vida em Nível Molecular*. [S.l.]: Artmed Editora, 2014.

2 NELSON, D. L.; COX, M. M. *Princípios de bioquímica de Lehninger*. [S.l.]: Artmed Editora, 2022.

3 CAVALCANTI, A. R.; LANDWEBER, L. F. Genetic code: what nature missed. *Current Biology*, Elsevier, v. 13, n. 22, p. R884–R885, 2003.

4 WAY, J. C. et al. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell*, Elsevier, v. 157, n. 1, p. 151–161, 2014.

5 ABOLA, E. E.; BERNSTEIN, F. C.; KOETZLE, T. F. The protein data bank. In: *neutrons in biology*. [S.l.]: Springer, 1984. p. 441–441.

6 LEE, Y. et al. Importance of protein dynamics in the structure-based drug discovery of class ag protein-coupled receptors (gpcrs). *Current opinion in structural biology*, Elsevier, v. 55, p. 147–153, 2019.

7 KLINMAN, J. P. Importance of protein dynamics during enzymatic c–h bond cleavage catalysis. *Biochemistry*, ACS Publications, v. 52, n. 12, p. 2068–2077, 2013.

8 MANDELKOW, E.; MANDELKOW, E.-M. Kinesin motors and disease. *Trends in cell biology*, Elsevier, v. 12, n. 12, p. 585–591, 2002.

9 KOLOMEISKY, A. B. Motor proteins and molecular motors: how to operate machines at the nanoscale. *Journal of Physics: Condensed Matter*, IOP Publishing, v. 25, n. 46, p. 463101, 2013.

10 KOLOMEISKY, A. B. *Motor proteins and molecular motors*. [S.l.]: CRC press, 2015.

11 OLDFIELD, C. J.; DUNKER, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*, v. 83, n. 1, p. 553–584, 2014.

12 DWEK, R. A. Glycobiology: toward understanding the function of sugars. *Chemical reviews*, ACS Publications, v. 96, n. 2, p. 683–720, 1996.

13 ARMANT, D. R.; KAPLAN, H. A.; LENNARZ, W. J. N-linked glycoprotein biosynthesis in the developing mouse embryo. *Developmental biology*, Elsevier, v. 113, n. 1, p. 228–237, 1986.

14 DOBSON, C. M. Protein folding and misfolding. *Nature*, Nature Publishing Group, v. 426, n. 6968, p. 884–890, 2003.

15 JUMPER, J. et al. Highly accurate protein structure prediction with alphafold. *Nature*, Nature Publishing Group, v. 596, n. 7873, p. 583–589, 2021.

16 ANFINSEN, C. B. Principles that govern the folding of protein chains. *Science*, American Association for the Advancement of Science, v. 181, n. 4096, p. 223–230, 1973.

17 LEVINTHAL, C. Are there pathways for protein folding? *Journal de chimie physique*, EDP Sciences, v. 65, p. 44–45, 1968.

18 CHAN, H. S.; DILL, K. A. Compact polymers. *Macromolecules*, ACS Publications, v. 22, n. 12, p. 4559–4573, 1989.

19 BRYNGELSON, J. D.; WOLYNES, P. G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *The Journal of Physical Chemistry*, ACS Publications, v. 93, n. 19, p. 6902–6915, 1989.

20 BALDWIN, R. L. How does protein folding get started? *Trends in biochemical sciences*, Elsevier, v. 14, n. 7, p. 291–294, 1989.

21 BRYNGELSON, J. D.; WOLYNES, P. G. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of sciences*, National Acad Sciences, v. 84, n. 21, p. 7524–7528, 1987.

22 ONUCHIC, J. N.; LUTHEY-SCHULTEN, Z.; WOLYNES, P. G. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 48, n. 1, p. 545–600, 1997.

23 LEOPOLD, P. E.; MONTAL, M.; ONUCHIC, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 89, n. 18, p. 8721–8725, 1992.

24 DILL, K. A. et al. The protein folding problem. *Annual review of biophysics*, NIH Public Access, v. 37, p. 289, 2008.

25 THIRUMALAI, D. et al. Theoretical perspectives on protein folding. *arXiv preprint arXiv:1007.3029*, 2010.

26 WHITFORD, P. C. et al. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 75, n. 2, p. 430–441, 2009.

27 WHITFORD, P. C. et al. Conformational transitions of adenylate kinase: switching by cracking. *Journal of molecular biology*, Elsevier, v. 366, n. 5, p. 1661–1671, 2007.

28 BEST, R. B.; CHEN, Y.-G.; HUMMER, G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure*, Elsevier, v. 13, n. 12, p. 1755–1763, 2005.

29 ZUCKERMAN, D. M. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *The Journal of Physical Chemistry B*, ACS Publications, v. 108, n. 16, p. 5127–5137, 2004.

30 LU, Q.; WANG, J. Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *Journal of the American Chemical Society*, ACS Publications, v. 130, n. 14, p. 4772–4783, 2008.

31 TAKETOMI, H.; UEDA, Y.; GŌ, N. 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*, Wiley Online Library, v. 7, n. 6, p. 445–459, 1975.

32 TAKADA, S. Gō model revisited. *Biophysics and physicobiology*, The Biophysical Society of Japan, v. 16, p. 248–255, 2019.

33 NOEL, J. K.; ONUCHIC, J. N. The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules. In: *Computational modeling of biological systems*. [S.l.]: Springer, 2012. p. 31–54.

34 CLEMENTI, C.; NYMEYER, H.; ONUCHIC, J. N. 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*, Elsevier, v. 298, n. 5, p. 937–953, 2000.

35 SOBOLEV, V. et al. Molecular docking using surface complementarity. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 25, n. 1, p. 120–129, 1996.

36 NOEL, J. K. et al. Smog 2: a versatile software package for generating structure-based models. *PLoS computational biology*, Public Library of Science San Francisco, CA USA, v. 12, n. 3, p. e1004794, 2016.

37 SHOEMAKER, B. A.; WANG, J.; WOLYNES, P. G. Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *Journal of molecular biology*, Elsevier, v. 287, n. 3, p. 675–694, 1999.

38 FERSHT, A. R. Characterizing transition states in protein folding: an essential step in the puzzle. *Current opinion in structural biology*, Elsevier, v. 5, n. 1, p. 79–84, 1995.

39 OLIVEIRA, L. C. d. Aspectos de topologia e mutação no processo de enovelamento e evolução de proteínas. Universidade Estadual Paulista (UNESP), 2008.

40 CHAVEZ, L. L.; ONUCHIC, J. N.; CLEMENTI, C. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *Journal of the American Chemical Society*, ACS Publications, v. 126, n. 27, p. 8426–8432, 2004.

41 PLOTKIN, S. S.; ONUCHIC, J. N. Investigation of routes and funnels in protein folding by free energy functional methods. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 97, n. 12, p. 6509–6514, 2000.

42 JR, A. B. O. et al. Distinguishing biomolecular pathways and metastable states. *Journal of chemical theory and computation*, ACS Publications, v. 15, n. 11, p. 6482–6490, 2019.

43 JUNIOR, A. B. O. et al. Exploring energy landscapes of intrinsically disordered proteins: Insights into functional mechanisms. *Journal of Chemical Theory and Computation*, ACS Publications, 2021.

44 HARDIN, C. et al. Associative memory hamiltonians for structure prediction without homology: alpha-helical proteins. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 97, n. 26, p. 14235–14240, 2000.

45 TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multi-dimensional data sets. *Information Visualization*, SAGE Publications Sage UK: London, England, v. 2, n. 4, p. 218–231, 2003.

46 PRONK, S. et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, Oxford University Press, v. 29, n. 7, p. 845–854, 2013.

47 LEMAK, A.; BALABAEV, N. On the berendsen thermostat. *Molecular Simulation*, Taylor & Francis, v. 13, n. 3, p. 177–187, 1994.

48 MARENGERE, L. E.; PAWSON, T. Structure and function of sh2 domains. *Journal of Cell Science*, Company of Biologists, v. 1994, n. Supplement_18, p. 97–104, 1994.

49 SANCHES, K. et al. Nmr assignment of free 1h, 15n and 13c-grb2-sh2 domain. *Biomolecular NMR Assignments*, Springer, v. 13, n. 2, p. 295–298, 2019.

50 KAWATA, T. et al. Sh2 signaling in a lower eukaryote: a stat protein that regulates stalk cell differentiation in dictyostelium. *Cell*, Elsevier, v. 89, n. 6, p. 909–916, 1997.

51 PAWSON, T.; NASH, P. Assembly of cell regulatory systems through protein interaction domains. *science*, American Association for the Advancement of Science, v. 300, n. 5618, p. 445–452, 2003.

52 MUSACCHIO, A.; WILMANNS, M.; SARASTER, M. Structure and function of the sh3 domain. *Progress in biophysics and molecular biology*, Elsevier, v. 61, n. 3, p. 283–297, 1994.

53 KANEKO, T.; LI, L.; LI, S. S. The sh3 domain–a family of versatile peptide-and protein-recognition module. *Frontiers in Bioscience-Landmark*, IMR Press, v. 13, n. 13, p. 4938–4952, 2008.

54 MAYER, B. J. The discovery of modular binding domains: building blocks of cell signalling. *Nature reviews Molecular cell biology*, Nature Publishing Group, v. 16, n. 11, p. 691–698, 2015.

55 DALGARNO, D.; BOTFIELD, M.; RICKLES, R. Sh3 domains and drug design: ligands, structure, and biological function. *Peptide Science*, Wiley Online Library, v. 43, n. 5, p. 383–400, 1997.

56 REN, R. et al. Identification of a ten-amino acid proline-rich sh3 binding site. *Science*, American Association for the Advancement of Science, v. 259, n. 5098, p. 1157–1161, 1993.

57 GIUBELLINO, A.; BURKE, T. R.; BOTTARO, D. P. Grb2 signaling in cell motility and cancer. *Expert opinion on therapeutic targets*, Taylor & Francis, v. 12, n. 8, p. 1021–1033, 2008.

58 ZHANG, S. et al. The role of the grb2–p38 mapk signaling pathway in cardiac hypertrophy and fibrosis. *The Journal of clinical investigation*, Am Soc Clin Investig, v. 111, n. 6, p. 833–841, 2003.

59 LIN, C.-C. et al. Inhibition of basal fgf receptor signaling by dimeric grb2. *Cell*, Elsevier, v. 149, n. 7, p. 1514–1524, 2012.

60 AHMED, Z. et al. Grb2 controls phosphorylation of fgfr2 by inhibiting receptor kinase and shp2 phosphatase activity. *Journal of Cell Biology*, The Rockefeller University Press, v. 200, n. 4, p. 493–504, 2013.

61 AHMED, Z. et al. Grb2 monomer–dimer equilibrium determines normal versus oncogenic function. *Nature communications*, Nature Publishing Group, v. 6, n. 1, p. 1–11, 2015.

62 SANCHES, K. et al. The dynamics of free and phosphopeptide-bound grb2-sh2 reveals two dynamically independent subdomains and an encounter complex with fuzzy interactions. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–13, 2020.

63 OCHI, S.; IIYAMA, M.; ODA, M. Interdomain interactions in grb2 revealed by the conformational stability and cd28 binding analysis. *Biophysical Chemistry*, Elsevier, v. 284, p. 106792, 2022.

64 HOU, B. et al. Grb2 binds to pten and regulates its nuclear translocation to maintain the genomic stability in dna damage response. *Cell death & disease*, Nature Publishing Group, v. 10, n. 8, p. 1–14, 2019.

65 YE, Z. et al. Grb2 enforces homology-directed repair initiation by mre11. *Science Advances*, American Association for the Advancement of Science, v. 7, n. 32, p. eabe9254, 2021.

66 SPIRO, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, Oxford University Press, v. 12, n. 4, p. 43R–56R, 2002.

67 SCHMIDT, M. A.; RILEY, L. W.; BENZ, I. Sweet new world: glycoproteins in bacterial pathogens. *Trends in microbiology*, Elsevier, v. 11, n. 12, p. 554–561, 2003.

68 ABU-QARN, M.; EICHLER, J.; SHARON, N. Not just for eukarya anymore: protein glycosylation in bacteria and archaea. *Current opinion in structural biology*, Elsevier, v. 18, n. 5, p. 544–550, 2008.

69 MITRA, N.; SHARON, N.; SUROLIA, A. Role of n-linked glycan in the unfolding pathway of erythrina corallodendron lectin. *Biochemistry*, ACS Publications, v. 42, n. 42, p. 12208–12216, 2003.

70 KAUSHIK, S.; MOHANTY, D.; SUROLIA, A. Role of glycosylation in structure and stability of erythrina corallodendron lectin (ecorl): a molecular dynamics study. *Protein Science*, Wiley Online Library, v. 20, n. 3, p. 465–481, 2011.

71 DESSEN, A. et al. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group i carbohydrate antigen. *Biochemistry*, ACS Publications, v. 34, n. 15, p. 4933–4942, 1995.

72 SINHA, S.; SUROLIA, A. Attributes of glycosylation in the establishment of the unfolding pathway of soybean agglutinin. *Biophysical journal*, Elsevier, v. 92, n. 1, p. 208–216, 2007.

73 HALDER, S.; SUROLIA, A.; MUKHOPADHYAY, C. Impact of glycosylation on stability, structure and unfolding of soybean agglutinin (sba): an insight from thermal perturbation molecular dynamics simulations. *Glycoconjugate journal*, Springer, v. 32, n. 6, p. 371–384, 2015.

74 LAPID, K.; SHARON, N. Meet the multifunctional and sexy glycoforms of glycodelin. *Glycobiology*, Oxford University Press, v. 16, n. 3, p. 39R–45R, 2006.

75 JOAO, H. C.; DWEK, R. A. Effects of glycosylation on protein structure and dynamics in ribonuclease b and some of its individual glycoforms. *European journal of biochemistry*, Wiley Online Library, v. 218, n. 1, p. 239–244, 1993.

76 BANKS, D. D. The effect of glycosylation on the folding kinetics of erythropoietin. *Journal of molecular biology*, Elsevier, v. 412, n. 3, p. 536–550, 2011.

77 RUBIO, M. V. et al. Redesigning n-glycosylation sites in a gh3 β-xylosidase improves the enzymatic efficiency. *Biotechnology for biofuels*, Springer, v. 12, n. 1, p. 1–14, 2019.

78 DUAN, L. et al. The sars-cov-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: implications for the design of spike-based vaccine immunogens. *Frontiers in immunology*, Frontiers Media SA, v. 11, p. 576622, 2020.

79 WITHKA, J. M. et al. Structure of the glycosylated adhesion domain of human t lymphocyte glycoprotein cd2. *Structure*, Elsevier, v. 1, n. 1, p. 69–81, 1993.

80 DAVIS, S. J.; MERWE, P. A. V. D. The structure and ligand interactions of cd2: implications for t-cell function. *Immunology today*, Elsevier, v. 17, n. 4, p. 177–187, 1996.

81 MIMURA, Y. et al. The influence of glycosylation on the thermal stability and effector function expression of human igg1-fc: properties of a series of truncated glycoforms. *Molecular immunology*, Elsevier, v. 37, n. 12-13, p. 697–706, 2000.

82 KIYOSHI, M. et al. Glycosylation of igg-fc: a molecular perspective. *International immunology*, Oxford University Press UK, v. 29, n. 7, p. 311–317, 2017.

83 ZHENG, K.; BANTOG, C.; BAYER, R. The impact of glycosylation on monoclonal antibody conformation and stability. In: TAYLOR & FRANCIS. *MAbs*. [S.l.], 2011. v. 3, n. 6, p. 568–576.

84 HULST, M. M.; MOORMANN, R. J. Erns protein of pestiviruses. *Methods in enzymology*, Elsevier, v. 342, p. 431–440, 2001.

85 SAINZ, I. F. et al. Removal of a n-linked glycosylation site of classical swine fever virus strain brescia erns glycoprotein affects virulence in swine. *Virology*, Elsevier, v. 370, n. 1, p. 122–129, 2008.

86 BRANZA-NICHITA, N. et al. Role of n-glycan trimming in the folding and secretion of the pestivirus protein erns. *Biochemical and biophysical research communications*, Elsevier, v. 319, n. 2, p. 655–662, 2004.

A Trabalhos publicados no período do doutorado

Heliyon 5 (2019) e02869

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Heliyon

Research article

Grb2 dimer interacts with Coumarin through SH2 domains: A combined experimental and molecular modeling study

Karoline Sanches [a,b], Raphael Vinicius Rodrigues Dias [a], Paulo Henrique da Silva [a], Marcelo Andrés Fossey [a,b], Ícaro Putinhon Caruso [a,b], Fátima Pereira de Souza [a,b], Leandro Cristante de Oliveira [a], Fernando Alves de Melo [a,b,*]

[a] Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University "Júlio de Mesquita Filho" (UNESP), São José do Rio Preto, SP, 15054-000, Brazil

[b] Multiuser Center for Biomolecular Innovation (CMIB), Institute of Biosciences, Humanities and Exact Sciences (IBILCE), São Paulo State University "Júlio de Mesquita Filho" (UNESP), São José do Rio Preto, SP, 15054-000, Brazil

ARTICLE INFO

Keywords:
Biochemistry
Molecular biology
Biophysics
Biophysical chemistry
Spectroscopy
Biomolecules
Molecular docking
Molecular dynamics
Grb2
Coumarin
Fluorescence
STD-NMR
Molecular docking

ABSTRACT

Grb2 is an important regulator of normal vs. oncogenic cell signaling transduction. It plays a pivotal role on kinase-mediated signaling transduction by linking Receptor Tyrosine kinases to Ras/MAPK pathway which is known to bring oncogenic outcome. Coumarins are phenolic molecules found in several plants and seeds widely studied because of the antibiotic, anti-inflammatory, anticoagulant, vasodilator, and anti-tumor properties. Despite several studies about the anti-tumor properties of Coumarin in vivo and the role of Grb2 in signaling pathways related to cell proliferation, a molecular level investigation of the interaction between Grb2 and Coumarin is still missing. In this study, we performed a combined set of biophysical approaches to get insights on the interaction between Grb2 in a dimer state and Coumarin. Our results showed that Coumarin interacts with Grb2 dimer through its SH2 domain. The interaction is entropically driven, 1:1 molecular ratio and presents equilibrium constant of 10^5 M^{-1}. In fact, SH2 is a well-known domain and a versatile signaling module for drug targeting which has been reported to bind compounds that block Ras activation in vivo. Despite we don't know the biological role coming from interaction between Grb2-SH2 domain and Coumarin, it is clear that this molecule could work in the same way as a SH2 domain inhibitor in order to block the link of Receptor Tyrosine kinases to Ras/MAPK pathway.

1. Introduction

Growth factor receptor-bound protein 2 (Grb2) is an important regulator of normal vs. oncogenic cell signaling transduction [1, 2, 3, 4]. It plays a pivotal role on Fibroblast Growth Factor Receptor 2 (FGFR2) activity before extracellular stimuli by binding to it as a dimer to bring together two FGFR2 molecules in order to form a tetrameric complex 2:2 which is inhibitory for FGFR2 [5]. It also is fundamental on kinase-mediated signaling transduction by linking receptor tyrosine kinases to Ras/MAPK (mitogen-activated protein kinase) pathway which is known to bring oncogenic outcome. More recently, it was reported that the equilibrium of Grb2 monomer-dimer is determinant to its normal versus oncogenic function [6]. In that study, the authors state that Grb2 can only bind to the protein son-of-sevenless (SOS) and upregulate MAPK signaling as a monomer, but inhibits the process as a dimer [6]. Grb2 has no intrinsically enzymatic activity and it is composed of a central Src homology 2 (SH2) domain placed between two Src homology 3 (SH3) domains [7]. The SH2 domain binds to specific phosphotyrosine containing motifs on Receptor Tyrosine Kinases (RTKs) and auxiliary docking proteins [8, 9] like the insulin receptor substrates (IRSs) [5, 10, 11, 12]. On the other hand, SH3 domains are reported to recognize proline-rich motifs on several proteins including FGFR2 [13], mammalian homologue of the drosophila SOS upstream of the MAPK pathway, a guanine nucleotide exchange factor (GEF) which acts on Ras, a GTPase subfamily protein, which in turn triggers a MAP Kinase cascade and other signaling pathways [5, 12, 14, 15]. Aberrancies in the FGFR2 and MAP Kinase mediated signaling pathway are known to be involved in many human developmental pathologies and several types of cancer [14, 15, 16]. The statement of Grb2 is a regulator of those important human signaling pathways open up a new window for searching molecules with

* Corresponding author.
E-mail address: fernando.melo@unesp.br (F.A. de Melo).

https://doi.org/10.1016/j.heliyon.2019.e02869
Received 10 April 2019; Received in revised form 19 October 2019; Accepted 12 November 2019
2405-8440/© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

anti-tumor properties to be characterized the interaction with Grb2 as a target protein involved in the growth of malignant cells [6].

There are several molecules with such intrinsic anti-tumor properties [17, 18, 19, 20]; among those molecules, Coumarins represent an important class of molecules that are easily found in vegetables, fruits, seeds, nuts, coffee, tea and wine. Coumarin exhibits interesting pharmacological properties that have been explored in many applications such as: antibiotic, anti-inflammatory, anticoagulant, vasodilator, anti-tumor among others [21]. The 1,2-benzopyrone is the structurally simplest member of Coumarin family and it has been reported that the treatment using this molecule showed decrease of metastatic renal carcinoma cells [22], prostatic carcinoma cells [23], breast cancer cells [24], and lung carcinoma cells [25].

Despite several studies about the anti-tumor properties of Coumarin and the role of Grb2 in signaling pathways related to cell proliferation, a molecular level investigation of the interaction between Grb2 and Coumarin is still missing. Here we present a study of the interaction between Grb2 in a dimer state and Coumarin by combining a biophysical multi-technique approach that includes Saturation Transfer Difference through Nuclear Magnetic Resonance (STD-NMR) and fluorescence experiments, molecular docking and molecular dynamics simulation. Our analysis showed that Coumarin can interact in a hydrophobic cavity placed at SH2 domain of Grb2 dimer (one per monomer). The role of this interaction is still unknown but this pocket on Grb2-SH2 domain is known to bind phosphorylated tyrosine residues on RTKs and their auxiliary docking proteins. Consequently Coumarin could somehow abrogate Grb2 interaction on cell signaling pathway mediated by RTKs in order to inhibit them.

2. Results

2.1. Binding constant and thermodynamic profile

Decrease in the fluorescence intensity can occur by several mechanisms, such as reactions with the excited state, molecular rearrangement, energy transfer, ground state complex formation, and collisional quenching [26]. The Fig. 1A shows the fluorescence emission spectra at 325 nm for Grb2 in a presence of Coumarin. Fluorescence intensity decreases while increasing Coumarin concentration in the solution, suggesting this molecule affects the micro-environment of tryptophan residues on Grb2 [26].

Grb2 concentration was set to 2 μM, which according to Lin and co-authors [5], Grb2 takes a dimer form. The maximum emission wavelength underwent a red shift up to 10 nm with the increment of Coumarin concentration. This red shift from 325 to 335 nm indicates that the polarity around the tryptophan residues involved in the interaction increases, suggesting a conformational change induced by ligand binding [26, 27]. Hence, these observations indicate that Coumarin interaction may occur in a protein region containing one or more tryptophans moderately exposed to the solvent. In order to determine the association constant for the interaction between Coumarin and Grb2, it was supposed as approach a model of two states as following [28, 29]:

$$P_f + L_f \rightarrow PL \tag{1}$$

where P_f, L_f states for free protein and free ligand in the solution, respectively, while PL states for bound protein. Assuming this model, protein-ligand association constant was calculated according to the following equation:

$$K_b = \frac{[PL]}{[P_f][L_f]} \tag{2}$$

where [PL] states for concentration of the protein-ligand complex, $[P_f]$ and $[L_f]$ states for free protein and free ligand concentrations in the solution, respectively. Rewriting [PL] to be $[P_b]$ (bound protein) and assuming that one ligand can bind to n sites of the protein, 1L → nPT (PT states for total protein concentration in the solution), it can use Scatchard equation [29] to determine K_b as following:

$$(n - \nu)K_b = \frac{\nu}{[L_f]} \tag{3}$$

where $\nu = \frac{[P_b]}{[P_T]}$ is the bound protein fraction, n is the number of ligand per protein and K_b is the association constant [30]. Scatchard plot is shown in Fig. 1B where it is possible to see a linear behavior which is characteristic for a non-cooperative binding process [26, 28]. Once K_b for each

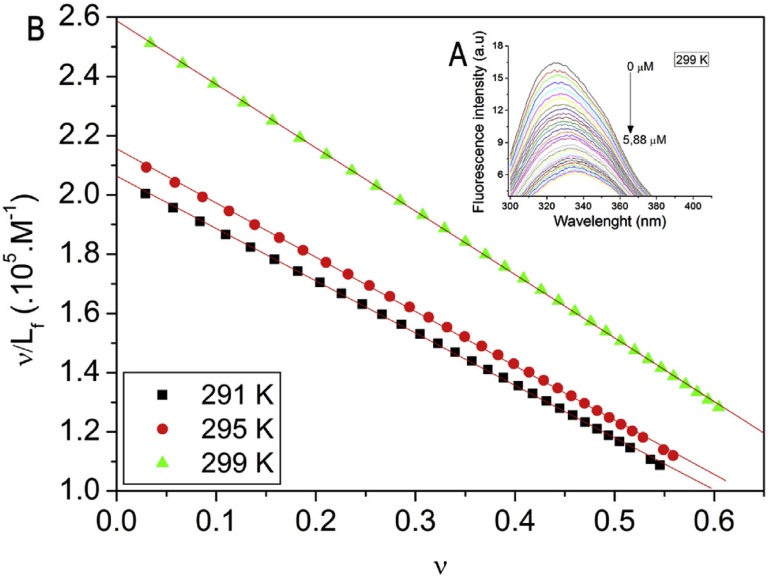


Fig. 1. (A) Emission spectra of the fluorescence intensity of Grb2 in absence and presence of increasing amount of Coumarin concentrations (from 0 to 5.88 μM). (B) Scatchard plots for the fluorescence quenching of Grb2 by the Coumarin ligand at pH 8.0 and 291 (black square), 295 (red circle) and 299K (green triangle) where K_b constant for Grb2-Coumarin interaction was calculated to be on the order of 10^5 M^{-1} assuming a 1:1 stoichiometry (n).

experimental temperature was obtained by Scatchard plot (Fig. 1B), thermodynamic profile of the Grb2-Coumarin interaction could be determined through the van't Hoff analysis (Fig. 2) by using the following equation [26]:

$$-\ln(K_b) = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \quad (4)$$

where R is the universal gas constant (R = 1.98 cal.K^{-1}mol^{-1}), ΔH and ΔS are the enthalpy and entropy changes, respectively. Results for association constants, number of ligands per protein and thermodynamic parameters of interaction can be found in Table 1.

Fluorescence quenching allows the verification of the binding forces involved in protein-ligand interaction by monitoring changes of thermodynamic parameters (i.e., enthalpy change ΔH and entropy change ΔS) [26, 31, 32].

Interaction between Grb2 dimer and Coumarin was found to be favorable as Gibbs free energy decreases while increasing the experimental temperature, and entropic driven as the term TΔS is the major contribution in the thermodynamic profile of Grb2-Coumarin binding (Fig. 2B). The stoichiometry was found to be 1:1 and the association constants determined at three different temperatures were on the order of 10^5 M^{-1}, indicating a moderate interaction between Grb2 and Coumarin (Table 1).

2.2. STD-NMR spectroscopy

STD-NMR (Saturation Transfer Difference based in Nuclear Magnetic Resonance) is a powerful method for epitope mapping. This method consists in the observation of the changes in saturation of ligand specifically bound to the target protein, through the selective saturation of the protein [33, 34, 35]. Hence, based on NOE (Nuclear Overhauser effect) and changes in ligand resonance signals, STD-NMR technique allows the study of protein-ligand interaction which dissociation constants (K_D) are in a range that goes from 10^{-8} M to 10^{-3} M. The main information obtained from this technique is the ligand epitope interaction map that can be calculated through STD amplification factor by using the Eq. (5) [35, 36, 37, 38].

Table 1
Binding and thermodynamic parameters obtained from Scatchard model and van't Hoff analysis for the Grb2-Coumarin interaction at pH 8.0 and 291, 295, and 299 K. This interaction is favorable, entropycally driven within a moderate association constant (10^5 M^{-1}) and the number of ligands by protein is equal to 1. Those parameters suggest that hydrophobic effects are the major contribution on the interaction of Grb2 and Coumarin.

T (K)	Kb (10^5 M^{-1})	n	ΔG (kcal.mol^{-1})	ΔH (kcal.mol^{-1})	TΔS (kcal.mol^{-1})
291	1.7	1.0	-6.95	3.84	-10.8
295	1.8	1.1	-7.07		-10.9
299	2.1	1.2	-7.25		-11.1

$$A_{STD} = \frac{I - I_{SAT}}{I}\frac{[L_T]}{[P]} = \frac{I_{STD}}{I}\frac{[L_T]}{[P]} \quad (5)$$

where I is the signal intensity after off-resonance saturation, I_{SAT} is the intensity of on-resonance saturated signal, I_{STD} signal is the difference I-I_{SAT}, $[L_T]$ is the ligand concentration and [P] the protein concentration.

The STD effect was verified through the intensity of each Coumarin atom from the reference to difference spectra and the epitope mapping was obtaining as proportional to the STD amplification factor. The reference spectra and the difference spectra are shown in Fig. 3 in red and blue, respectively, and the results thus obtained are shown in Table 2.

The assignment ^{1}H NMR spectrum for Coumarin was done by consulting the SDBS (Spectral Database for Organic Compounds) and can be seen in the red spectrum (Fig. 3). From the epitope mapping it can be observed that the Coumarin hydrogens are surrounded by amino acid residues in the pocket of Grb2 protein when the interaction happens. Another information obtained from this approach is the spatial orientation of Coumarin in the binding site which according to the epitope mapping the hydrogens 5 and 8 are closer to the binding site in the protein, whereas these protons received more saturation transfer (Table 2). This result suggests that the benzene ring of Coumarin could be buried in the protein binding site.

2.3. Molecular docking and molecular dynamics simulations

A combined analysis of fluorescence and NMR results revealed that

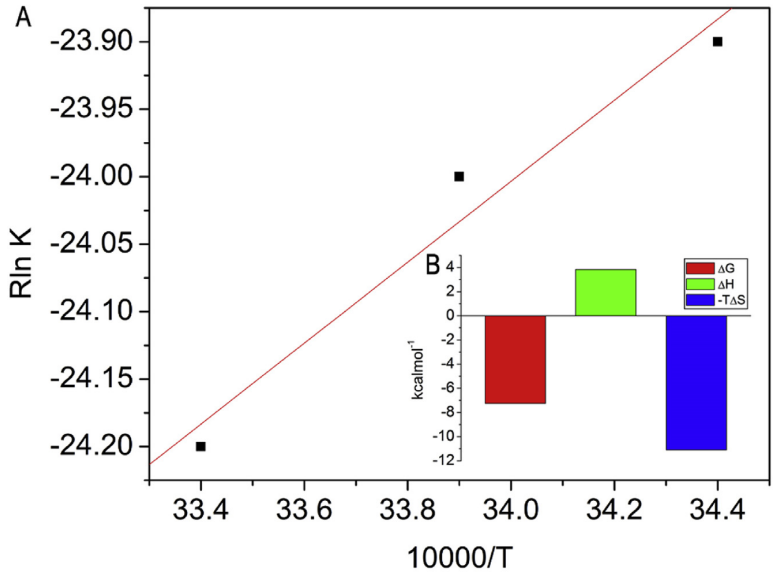


Fig. 2. (A) Van't Hoff plot and (B) thermodynamic profile showing the interaction to be spontaneous (ΔG < 0) and entropically driven (TΔS > 0) while enthalpy counts unfavorable to the interaction (ΔH > 0). This behavior is frequently associated to hydrophobic effects taking place as the major interaction mechanism [56].

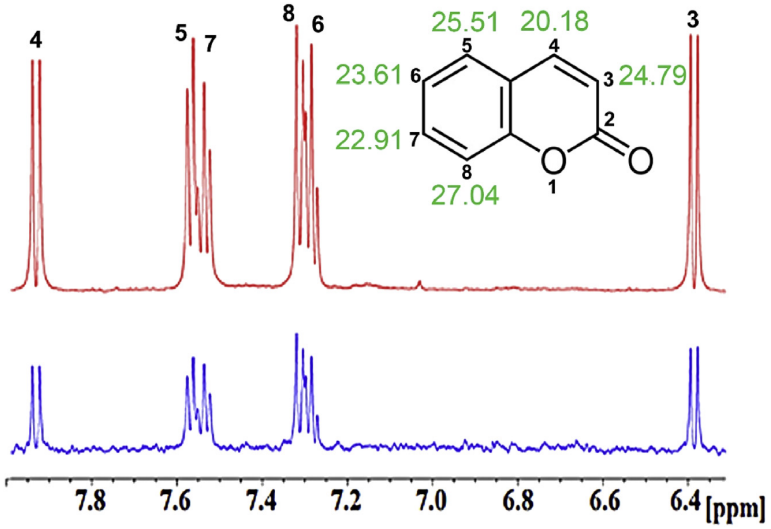


Fig. 3. Reference ^{1}H NMR of Coumarin (red) and I_{STD} spectra (blue). The signals in the I_{STD} due the saturation transfer from Grb2 to Coumarin reveals the interaction between these molecules. The STD amplification factor indicates that Coumarin is surrounded by Grb2 amino acids residues, where hydrogen atoms 5 and 8 are receiving the highest saturation transfer.

Table 2
Hydrogen position, 1D ^{1}H-NMR Coumarin chemical shift, I_{STD} and STD amplification factor. It is shown that Coumarin interacts with Grb2 in a way that hydrogen atoms 5 and 8 are the ones receiving more saturation transfer.

^{1}H	δ (ppm)	I_{STD}	A_{STD}
3	6.38	0.0925	24.79
4	7.9	0.0753	20.18
5	7.53	0.0952	25.51
6	7.28	0.0881	23.61
7	7.56	0.0855	22.91
8	7.31	0.1009	27.04

Coumarin interacts with a hydrophobic region of Grb2 containing one or more tryptophan residues. Thus, a search for the interacting region was done through molecular docking. Initially, the four largest superficial cavities were considered possible sites for this interaction and it was performed with the protein in its dimeric form (Fig. 4A) in order to match *in vitro* experimental insights with *in silico* calculations.

The search indicated four possible pockets for the ligand. However, only one was suitable with the experimental results for the interaction (Fig. 4B) located in SH2 domain of Grb2. Thus molecular docking calculations were performed using the SH2 domain and directed to the selected cavity. For statistical proposes 1500 poses were generated. The most representative conformation was selected using AMBER Score Energy and pose frequency [39] (Fig. 4C). Calculations were done considering the conformation with the lowest AMBER Score Binding Energy (as reference value) and comparing the Root Mean Square Deviation (RMSD) against all other structures. The RMSD measures the similarity between two structures and indicated 92.8% of the selected conformations with values between 0 and 1 Å (excellent agreement). Differences between 1 and 2 Å in the RMSD values were not found, 6% was found between 2 and 3 Å and only 1.2% of the values were superior to 3 Å (Fig. 5C). Those results have shown a relationship between Coumarin accessibility to different protein conformations and the interaction energy.

Since the molecular docking method used here is severely restrictive, not allowing hydration effects or side chain/ligand accommodation, was performed a molecular dynamics simulations by 20 ns using GROMACS Package [40]. The trajectories obtained from MD simulations revealed conformational changes in the protein surface, caused mainly by the movements of the side chains of the residues Ser90, Ala115, Trp121 and Arg142. These changes allow the migration of Coumarin to a more buried region of the selected pocket, resulting in the stability of the hydrophobic interaction.

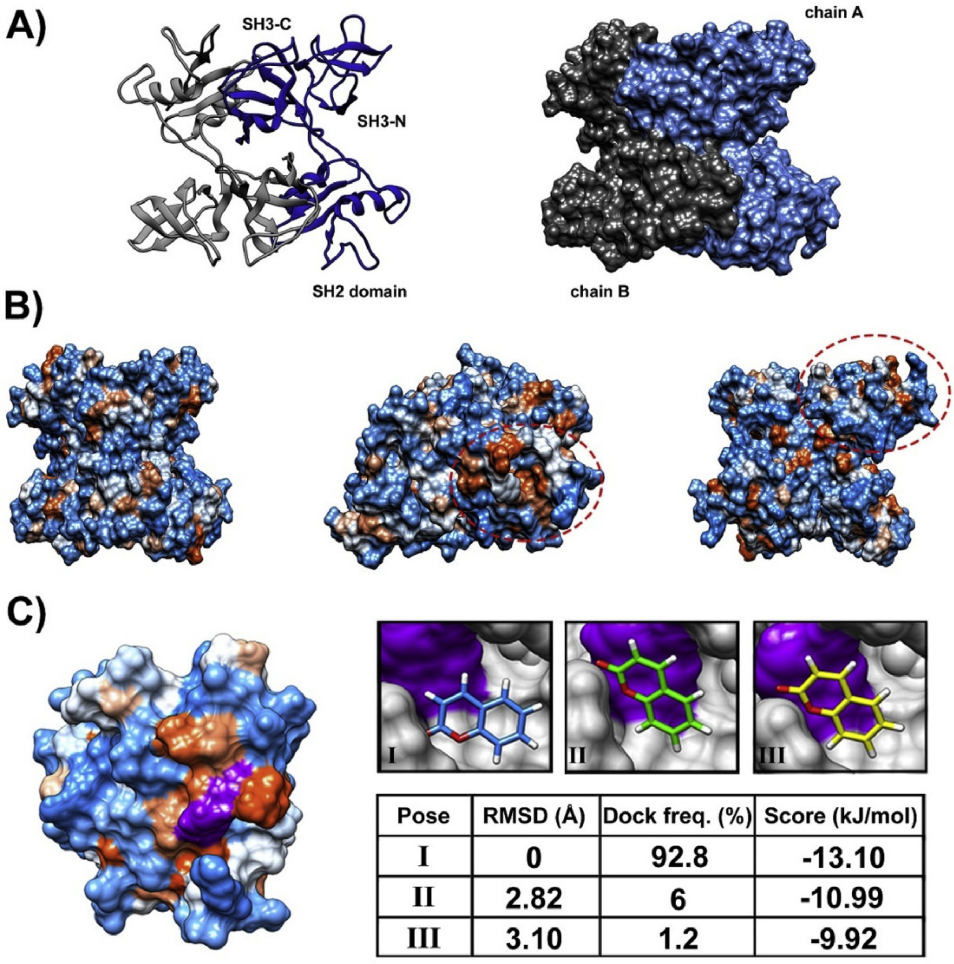
Molecular dynamics trajectories from simulations were analyzed considering the energy average of short-range interactions between the protein and the ligand as comparison parameter, based on CHARMM27 force field [41]. The best conformations were selected using the lowest value of Van der Waals energy (vdw). Then, to keep an accurate energetic comparison, the complexed structure as separated in receptor and ligand, and resubmitted to the entire docking process.

The analysis showed a final value of the interaction energy obtained with AmberScore by DOCK6 of -20.62 kJ/mol. That is a lower energy value when compared to the three previous energies obtained before molecular dynamics (-13.10, -10.99, and 9.92 kJ/mol) presented in Fig. 4. Also, a more detailed analysis of the docking region has indicated Coumarin orients itself in a way that carbons are surrounded by the hydrophobic residues in from Grb2-SH2 domain pocket while oxygen atoms are faced to solvent (Fig. 5B).

The calculations indicated contacts between the Coumarin and the residues Trp121, Val88, Phe108, Leu97, His135, and Val40 (Figs. 5B and 5D) using distances less than 4 Å as the criterion of analysis. A detailed representation of the interactions can be visualized by using LIGPLOT software [42] which has revealed these six amino acids residues inside Grb2-SH2 domain pocket to hydrophobic interact with Coumarin. Arrangements of Coumarin rings with the residues Trp121 and His135 have shown a stacking-like interaction by PLIP calculations and it was included in the figure. Further analysis was performed using the MM-PBSA method [43] in which was identified the residues with highest energy contribution to the protein-ligand interaction. The result is corroborated with previous results of molecular docking, highlighting an interaction with Trp121 (Fig. 5A). Additionally, two water molecules are taking part in the interaction assembly by forming hydrogen bonds with the oxygen bound in the carbonyl of Coumarin. This interaction was calculated using the h_bond tool [44], resulting an average of 1.7 hydrogen bonds between ligand and waters during the entire trajectory.

3. Discussion

Grb2 is an important adaptor protein which is ubiquitously expressed



Pose	RMSD (Å)	Dock freq. (%)	Score (kJ/mol)
I	0	92.8	-13.10
II	2.82	6	-10.99
III	3.10	1.2	-9.92

Fig. 4. (A) Ribbons representation of the dimeric structure of the Grb2 protein used for computational analysis (PDBid: 1GRI). Monomers are represented in ribbons (left) and surface (right). They are distinguished in blue and gray for the chains A and B, respectively. (B) Surface representation of the Grb2 structure colored by hydrophobicity in a blue to orange scale, rotated 180° vertically twice. The SH2 domain is highlighted and presented in 3 different poses. The combined analysis of these initial attempts with the experimental results indicates only one region with reasonable accessibility, residues with similar hydrophobicity condition and presence of a tryptophan. (C) Grb2-SH2 domain surface representation employed for molecular docking calculations. The tryptophan residue is highlighted in purple and localized in the bottom of pocket (from this perspective view). This is the conformation used for Coumarin molecular docking calculations. Each pose obtained was clustered according to its RMSD and separated into three different groups, showing that the group with the highest access number (92.8%) it is also the ensemble with the lowest computational energies (around -13.10 kJ/mol).

in eukaryotic cells. It is involved in many intracellular protein interactions and readily detectable in malignant forms of breast, prostate, and colon cancers, for example. Grb2 interacts as a dimer with FGFR2 kinase and regulates its activity before extracellular stimuli [5]. Also as a dimer, it links Receptor Tyrosine Kinases signaling to Ras/MAPK pathway and regulates MAPK pathway activity through its monomer-dimer equilibrium [6]. On the other hand, Coumarins are a class of molecules presenting anti-tumor properties [17, 18, 19] which has shown efficiency on the treatment of several types of cancer cells including prostatic and breast cancer [23, 24]. But there is also a lack of information in a molecular level about them that could lead to a more efficient drug development for cancer treatment if taking Grb2 as a protein target. So, in order to characterize the interaction between Grb2 and Coumarin we have used a combination of experimental and computational analysis which started with fluorescence spectroscopy as following. Fluorescence quenching has shown the interaction between Coumarin and Grb2 dimer to be moderate with binding constant on order of $10^5 M^{-1}$ and stoichiometry 1:1 according to Scatchard Model [29]. The van't Hoff analysis have revealed interaction to be favorable as Gibbs free energy changing supports that increasing in affinity and stability are both directly proportional to the temperature increments. Raising the temperature causes a decrease in the system micro-viscosity while increasing the molecular collision due to faster diffusion which makes the interaction favorable. The calculated values for $\Delta G < 0$ are characteristic of a spontaneous reaction while $T\Delta S > 0$ indicates that entropy drives the interaction through hydrophobic effects (Fig. 2B).

The major contribution of entropy for ΔG occurs when water molecules are released from the protein solvation layer during interaction or due to protein conformational changes induced by the ligand [45]. The hydrophobic nature of Coumarin molecule also corroborates with the thermodynamic analysis. NMR-STD experiments confirmed interaction between Coumarin and Grb2. By mapping the interaction epitopes for Coumarin and Grb2, it gives us a clue of how Coumarin could dock on Grb2 binding site (Fig. 3). According to that, Coumarin is surrounded by amino acid residues during interaction with Grb2 dimer and the hydrogens 5 and 8 are receiving the highest saturation transfer (Fig. 5). Next

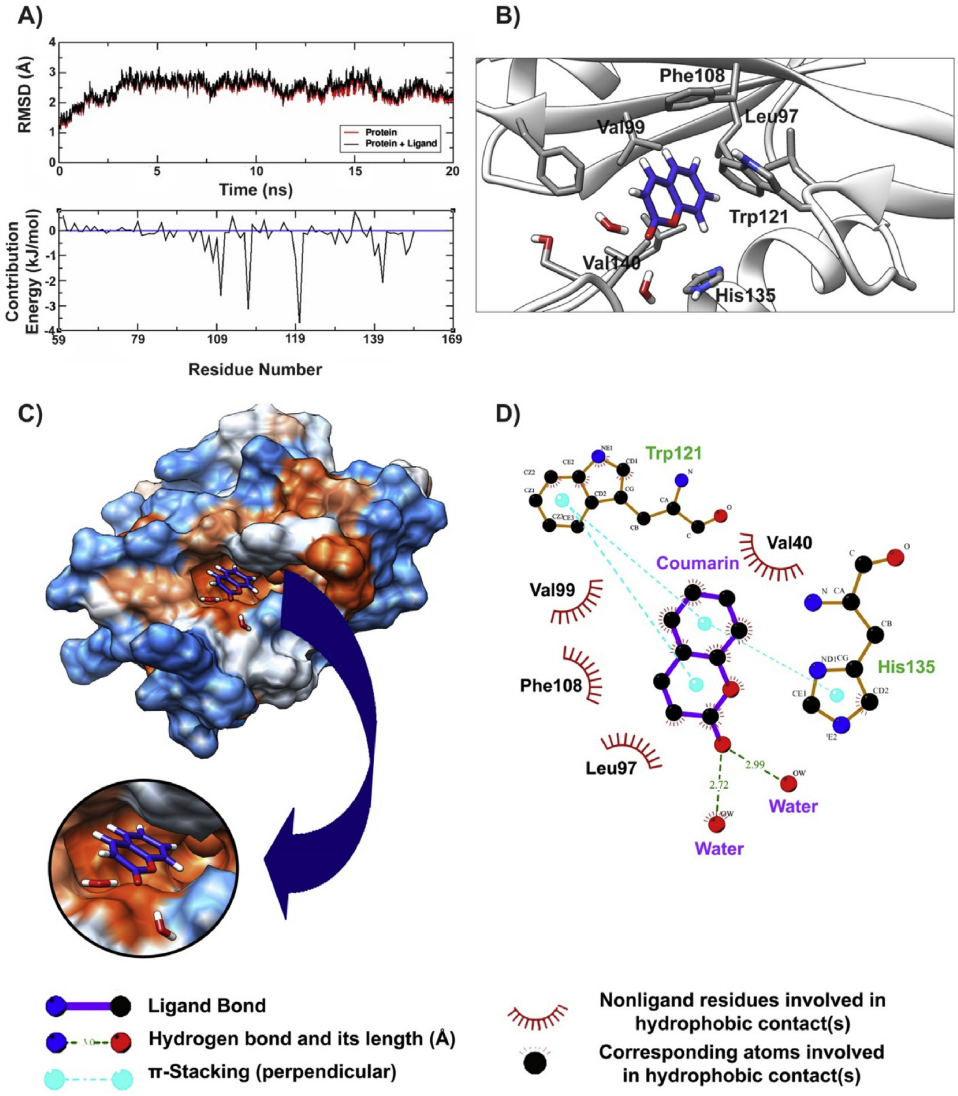


Fig. 5. Representation of the complex Grb2-Coumarin for the interaction in the Grb2-SH2 domain. (A) The upper graph is the RMSD of the protein (red) and the protein complexed with Coumarin (black) using as reference the first frame of the simulation that is the lower energy structure acquired from molecular docking. The bellow one is the energetic contribution for each residue calculated using the full molecular dynamics trajectory. For these calculations, we used the GROMACS package tools g_rms and g_mmpbsa [40, 43]. (B) Three-dimensional representation using cartoons and traces of the interaction pocket. Molecular representations were done using CHIMERA 1.7 suite [44]. (C) Grb2-SH2 domain surface representation colored by a hydrophobic scale from blue (lower) to orange (higher) with Coumarin docked in the pocket. Coumarin was colored with carbon atoms in blue, oxygen atoms in red, and hydrogen atoms in white. Water molecules taking place in the interactions are highlighted following the same patterns of nuclei. (D) Map of interactions of the complex Grb2-SH2-Coumarin. The map was obtained from two from different programs: LigPlot and PLIP. The representation was generated using LigPlot and stacking type interactions calculated by PLIP were added.

step was to employ computational simulations to determine the region on Grb2 dimer surface where Coumarin could be interacting with. In order to do that we search for a binding site which matches with experimental information as following: binding site should be hydrophobic containing one or more tryptophan residues in the bottom of it. Our computational approach found the SH2 domain on Grb2 dimer surface to be the best energy score for the interaction with Coumarin according to experimental data (Fig. 4). The binding site found in this domain is formed by hydrophobic residues and contains a tryptophan residue (Trp121) in the bottom of this cavity so in agreement with experimental data. Other places such as C-terminal Grb2-SH3 domain on the Grb2 dimer surface and Grb2 dimer interface were found which Coumarin has docked during the binding site computational search but none of than have totally matched to experimental information either by missing a tryptophan residue in cavity or by the hydrophilic nature of protein region. Besides, even Coumarin being capable of interacting with Grb2 domains separately (data not shown) by the same mechanism described here, our data suggests that Coumarin only binds to SH2 domain of Grb2 in its dimer form.

In fact, SH2 is a well-known domain and a versatile signaling module for drug targeting which has been reported to bind compounds that block Ras activation in vivo [46, 47, 48]. Despite we don't know the biological role coming from interaction between Grb2-SH2 domain and Coumarin it is clear that this molecule could work in the same way as SH2 domain inhibitors [46, 47, 48, 49] in order to block the link for Ras/MAPK signaling transduction as well. Lots of work still may be done to

understand the molecular basis which leads to cancer out come and exploring Grb2 biological role on all this process could be the way for that and for more efficient drugs development against the disease.

4. Materials and methods

4.1. Grb2 expression and purification

Grb2 protein - 6x histidine tagged - was expressed in *E. coli* BL21 (DE3) Gold. A single colony was used to transform 10 mL of LB medium which was grown for 15 h at 37 °C and shaking at 90 rpm. Then, this culture was used to inoculate 490 mL of LB containing 50 μg/mL kanamycin and 5 μg/mL tetracycline. This bacterial culture was let to shake at 37 °C until the $OD_{600} = 0.6$ being reached. After that, the temperature was lowered to 20 °C and culture was let to temperature equilibrate before induce expression by the addition of 0.4 mM IPTG. The culture was allowed to grow for a further 12 h before harvesting by centrifugation at 4000×g. The cells were re-suspended in buffer 50 mM Tris-HCl (pH 8.0) 300 mM NaCl, 1.0 mM β-mercaptoethanol (βME) in the presence protease inhibitors cocktail (Sigma Aldrich P2714-1BTL) and lysed by sonication. Cell drebis were separated by centrifugation (20,000×g at 4 °C for 45 min) and the soluble fraction was applied in metal affinity column IMAC HiTrap HP (GE Life Sciences) containing Co^{2+} ions. Following 50 mL buffer wash Grb2 was eluted with 200 mM imidazole in 50 mM Tris-HCl (pH 8.0) with 100 mM NaCl and 1.0 mM βME. The sample was concentrated to 1 mL and applied to a Superdex 75 10/300 (GE Life Sciences) gel filtration column in buffer containing 20 mM NaH_2PO_4 (pH 8.0) with 50 mM NaCl. Sample purity was confirmed by 15% SDS-PAGE to be higher than 98%.

4.2. Stock solution

The Coumarin (Sigma Aldrich) (MM 146.14 g/mol) stock solution was prepared in 90% (v/v) ethanol and kept on ice after that for up to 24 h before the interaction assays.

4.3. UV-Vis absorbance spectroscopy

UV-Vis measurements were performed on Thermo Scientific BIOMATE UV-Visible (Thermo Fisher Scientific) spectrophotometer equipped with a quartz cell of 1.0 cm path length, scanning speed of 600 nm/min, 1.0 nm of interval and spectral bandwidth of 2.0 nm. Spectra acquisition were performed at room temperature. Sample stock concentrations were obtained using the extinction coefficient of 38.055 and 11.438 $M^{-1}cm^{-1}$ for Grb2 and Coumarin, respectively [50, 51].

4.4. Fluorescence spectroscopy

Steady-state fluorescence spectroscopy measurements were performed on a spectrophotometer ISS PC1 (Champaign II, USA) equipped with a quartz cell of 1.0 cm path length and Nestlab RTE-221 thermostat bath (Thermo Electron Corporation, USA). Excitation wavelength at 290 nm was chosen since it provides excitation of tryptophan residues of protein. Emission spectrum was collected in the range of 300–500 nm, which was corrected for the background fluorescence of the buffer and for inner filter effects [26]. In the fluorescence quenching experiments performed at 291, 295, and 299 K, titrations were done by adding 2 μL from ligand stock solution (202 μM) to 2 μM Grb2 solution (2 mL) in buffer 20 mM NaH_2PO_4 (pH 8.0) with 50 mM NaCl. The final ligand concentration achieved with titrations was 5.88μM and the final ethanol concentration was 3% (v/v).

4.5. NMR spectroscopy

NMR experiments were performed on a Bruker Avance III 600.13 MHz equipped with a triple resonance 5 mm cryoprobe, having the pulse field gradient along the z axis. All data was analyzed with Bruker Topspin v3.2. Protein solution (5 μM of Grb2 in 20 mM NaH_2PO_4 buffer at pH 8.0, containing 50 mM NaCl) was used for determining the best saturation condition. It was tested at different on-resonance frequencies at -0.5, -1.0, and -1.5 ppm, keeping the off-resonance frequency at 20 ppm. Protein saturation time was reached at 3 s and recycle delay at 2 s. A total of 10 k scans were collected with 4 dummy scans. Saturation power was set to 1 ms and spinlock filter of 0.776 W was set to 1 ms at 288 K. When Coumarin was added to the system a spin lock filter of 30 ms was set for protein signals suppression. It was used 1.34 mM of Coumarin solubilized in ethanol-d6 CD_3CD_2OD (Sigma 186414).

4.6. Molecular docking and molecular dynamics simulations

Molecular docking simulations of Grb2-Coumarin complex were performed employing the tridimensional information deposited in the Protein Data Bank [52] (PDBid.: 1GRI) and the tridimensional structure of Coumarin from Compound Identification number (CID.: 323). Ligand and receptor were prepared using UCSF Chimera 1.7 [44]. Molecular docking calculations were done using UCSF Dock 6.7 package [53]. Protein cavities were identified using SPGHEN tool included in UCSF DOCK package. Throughout the docking process, the receptor Grb2 was set as rigid and the ligand as flexible initially (GRID Score step), followed by ligand and interaction receptor region flexible (AMBER rescoring). The grid box used for the ligand was divided in bins of 0.2 Å and the distance tolerance for matching ligand atoms to receptor was set to 0.75 Å. Computational docking calculation employed “anchor and grow” algorithm [54]. The first step of molecular docking was evaluated with single grid energy (SGE) score function composed by the van der Waals and the electrostatic interaction terms, also identified as Grid Score step. Then, a rescoring ranking was applied to the conformations obtained from SGE through Amber Score Binding Energy (ASBE). The calculation uses $E_{complex} = E_{binding} - (E_{receptor} - E_{ligand})$, where $E_{complex}$, $E_{receptor}$, and E_{ligand} are energies approximated by the Amber force field with MM-GB/SA method [53]. During this step small accommodations are allowed due the receptor and ligand flexibility.

Refinements of the Grb2-Coumarin complex were simulated through molecular dynamics, using as initial structure the conformation with the lowest AMBER score. The simulations were carried out using GROMACS package version 4.5.5 [40]. It was employed the force field CHARMM27 [41] with the standard set up for the protein residues and TIP3P water model. The ligand parameterization was performed using the webserver Swiss-param from Swiss Institute of Bioinformatics [55]. The MD minimization was done using 50000 steepest-descent steps and 5000 conjugate gradient steps, both without position restraints. The equilibration was performed in two steps with 100 ps, initially with and followed without position restrictions for the protein and ligand atoms. Initial atom velocities were based on Maxwell-Boltzmann distribution and the MD simulations for 20 ns with an integration step of 2 fs at 295 K, salt concentration of 0.15 M and 1 atm. The trajectory positions and energies were recorded each 100 ps. The simulation was performed using the Parrinello-Rahman barostat, Berendsen thermostat, the LYNCS method was used to constrain all hydrogen bonds and the simulation box was built with the edges of the box 10Å from the protein surface with Periodic Boundary Conditions (PBC). MD analyses were done using tools available in GROMACS package [40]. The conformation with the lowest energy of short range binding during all molecular dynamic simulation was selected, being considered the most probable conformation for the complex $Grb2_{dimeric}$-Coumarin. Since molecular simulations and docking calculations employs different energy expressions, the conformation with the lowest energy from simulations was submitted to a complete molecular docking procedure (SGE and ASBE), providing comparable energies.

Declarations

Author contribution statement

Karoline Sanches, Raphael Dias: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Paulo Silva: Performed the experiments; Analyzed and interpreted the data.

Marcelo Fossey, Fatima Souza, Fernando Alves de Melo: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Icaro Caruso, Leandro Oliveira: Conceived and designed the experiments; Analyzed and interpreted the data.

Funding statement

This work was supported by FAPESP grant no. 2009/53989-4, 2014/17630-0, 2016/08753-6 and CNPq grant no. 442951/2014-0 and 442352/2014-0.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors gratefully acknowledge the support of Dr. John E. Ladbury (University of Leeds, United Kingdom) for providing the Grb2 plasmid vector; Dr. Fabio R. de Moraes (Multiuser Center for Biomolecular Innovation, São Paulo State University, Brasil) for the support with the NMR data. We also thank Dr. Marcio F. Colombo (Department of Physics, São Paulo State University, Brasil) for support and for allowing us to work on the Spectroscopy Laboratory. The simulations were performed at the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) and CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em São Paulo), project UNICAMP/FINEP-MCTII.

References

[1] A.M. Tari, G. Lopez-Berestein, GRB2: a pivotal protein in signal transduction, Semin. Oncol. 28 (2001) 142–147.
[2] W. Kolch, M. Halasz, M. Granovskaya, B.N. Kholodenko, The dynamic control of signal transduction networks in cancer cells, Nat. Rev. Cancer 15 (2015) 515–527.
[3] A. Giubellino, T.R. Burke, D.P. Bottaro, Grb2 signaling in cell motility and cancer, Expert Opin. Ther. Targets 12 (2008) 1021–1033.
[4] E. Skolnik, A. Batzer, N. Li, C. Lee, E. Lowenstein, M. Mohammadi, B. Margolis, J. Schlessinger, The function of GRB2 in linking the insulin receptor to Ras signaling pathways, Science (80-) 260 (1993) 1953–1955.
[5] C.C. Lin, F.A. Melo, R. Ghosh, K.M. Suen, L.J. Stagg, J. Kirkpatrick, S.T. Arold, Z. Ahmed, J.E. Ladbury, Inhibition of basal FGF receptor signaling by dimeric Grb2, Cell 149 (2012) 1514–1524.
[6] Z. Ahmed, Z. Timsah, K.M. Suen, N.P. Cook, G.R. Lee, C.C. Lin, M. Gagea, A.A. Marti, J.E. Ladbury, Grb2 monomer-dimer equilibrium determines normal versus oncogenic function, Nat. Commun. 6 (2015) 1–9.
[7] S. Maignan, J.P. Guilloteau, N. Fromage, B. Arnoux, J. Becquart, A. Ducruix, Crystal structure of the mammalian Grb2 adaptor, Science (80-) 268 (1995) 291–293.
[8] E.J. Lowenstein, R.J. Daly, A.G. Batzer, W. Li, B. Margolis, R. Lammers, A. Ullrich, E.Y. Skolnik, D. Bar-Sagi, J. Schlessinger, The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling, Cell 70 (1992) 431–442.
[9] K. Sanches, Í.P. Caruso, F.C.L. Almeida, F.A. Melo, NMR assignment of free 1H, 15N and 13C-Grb2-SH2 domain, Biomol. NMR Assign. (2019).
[10] G. Gorgisen, I.M. Gulacar, O.N. Ozes, The role of insulin receptor substrate (IRS) proteins in oncogenic transformation, Cell. Mol. Biol. 63 (2017) 1–5.
[11] S.R. Hubbard, The insulin receptor: both a prototypical and atypical receptor tyrosine kinase, Cold Spring Harb. Perspect. Biol. 5 (2013) a008946.
[12] M. Rozakis-Adcock, R. Fernley, J. Wade, T. Pawson, D. Bowtell, The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1, Nature 363 (1993) 83–85.
[13] Z. Ahmed, C.C. Lin, K.M. Suen, F.A. Melo, J.A. Levitt, K. Suhling, J.E. Ladbury, Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity, J. Cell Biol. 200 (2013) 493–504.
[14] A.S. Dhillon, S. Hagan, O. Rath, W. Kolch, MAP kinase signalling pathways in cancer, Oncogene 26 (2007) 3279–3290.
[15] R.X.-D. Song, R.A. McPherson, L. Adam, Y. Bao, M. Shupnik, R. Kumar, R.J. Santen, Linkage of rapid estrogen action to MAPK activation by ERα-Shc association and Shc pathway activation, Mol. Endocrinol. 16 (2002) 116–127.
[16] I.S. Babina, N.C. Turner, Advances and challenges in targeting FGFR signalling in cancer, Nat. Rev. Cancer 17 (2017) 318–332.
[17] R. Miri, M. Nejati, L. Saso, F. Khakdan, B. Parshad, D. Mathur, V.S. Parmar, M.E. Bracke, A.K. Prasad, S.K. Sharma, O. Firuzi, Structure–activity relationship studies of 4-methylcoumarin derivatives as anticancer agents, Pharm. Biol. 54 (2016) 105–110.
[18] Y.A.-M. Abu BakarMohamad, Ahmed Al-Amiery, Abdul Amir Kadhum, Antioxidant activity of Coumarins 1 (2017) 24–30.
[19] R. Arora, S. Sawney, V. Saini, C. Steffi, M. Tiwari, D. Saluja, Esculetin induces antiproliferative and apoptotic response in pancreatic cancer cells by directly binding to KEAP1, Mol. Cancer 15 (2016) 64.
[20] V. Thomas, V. Thomas, D. Giles, D. Giles, G.P.M. Basavarajaswamy, G.P.M. Basavarajaswamy, A. Kumar Das, A. Kumar Das, A. Patel, A. Patel, Coumarin derivatives as anti-inflammatory and anticancer agents, Anti-Cancer Agents Med. Chem. Anti-Cancer Agents) 17 (2017) 415–423.
[21] S. Emami, S. Dadashpour, Current developments of coumarin-based anti-cancer agents in medicinal chemistry, Eur. J. Med. Chem. 102 (2015) 611–630.
[22] M.E. Marshall, L. Mendelsohn, K. Butler, L. Riley, J. Cantrell, C. Wiseman, R. Taylor, J.S. Macdonald, Treatment of metastatic renal cell carcinoma with coumarin (1,2-benzopyrone) and cimetidine: a pilot study, J. Clin. Oncol. 5 (1987) 862–866.
[23] J.L. Mohler, B.T. Williams, I.M. Thompson, M.E. Marshall, Coumarin (1,2-benzopyrone) for the treatment of prostatic carcinoma, J. Cancer Res. Clin. Oncol. 120 (Suppl) (1994). S35–38.
[24] M.A. Musa, J.S. Cooperwood, M.O.F. Khan, A review of coumarin derivatives in pharmacotherapy of breast cancer, Curr. Med. Chem. 15 (2008) 2664–2679. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3772644/.
[25] J.S. Lopez-Gonzalez, H. Prado-Garcia, D. Aguilar-Cazares, J.A. Molina-Guarneros, J. Morales-Fuentes, J.J. Mandoki, Apoptosis and cell cycle disturbances induced by coumarin and 7-hydroxycoumarin on human lung carcinoma cell lines, Lung Cancer 43 (2004) 275–283.
[26] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, third ed., Springer, New York, 2011.
[27] J.T. Vivian, P.R. Callis, Mechanisms of tryptophan fluorescence shifts in proteins, Biophys. J. 80 (2001) 2093–2109. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301402/.
[28] F. Ding, G. Zhao, J. Huang, Y. Sun, L. Zhang, Fluorescence spectroscopic investigation of the interaction between chloramphenicol and lysozyme, Eur. J. Med. Chem. 44 (2009) 4083–4089.
[29] F.W. Dahlquist, The meaning of Scatchard and Hill plots, Methods Enzymol. 48 (1978) 270–299.
[30] Charles R. Cantor, Paul R. Schimmel, Biophysical Chemistry - the Behavior of Biological Macromolecules, W. H. Freeman, 1971.
[31] C. Hao, G. Xu, Y. Feng, L. Lu, W. Sun, R. Sun, Fluorescence quenching study on the interaction of ferroferric oxide nanoparticles with bovine serum albumin, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 184 (2017) 191–197.
[32] A. Papadopoulou, R.J. Green, R.A. Frazier, Interaction of flavonoids with bovine serum albumin: a fluorescence quenching study, J. Agric. Food Chem. 53 (2005) 158–163.
[33] M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, J. Am. Chem. Soc. 123 (2001) 6108–6117.
[34] B. Meyer, T. Peters, NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, Angew. Chem. Int. Ed. 42 (2003) 864–890.
[35] A. Viegas, J. Manso, F.L. Nobrega, E.J. Cabrita, Saturation-transfer difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding, J. Chem. Educ. 88 (2011) 990–994.
[36] M. Meyer, B. Meyer, Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor, 2001. http://pubs.acs.org/doi/abs/10.1021/ja0100120.
[37] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, Angew. Chem. Int. Ed. 38 (1999) 1784–1788.
[38] Nissler Ludwig, Rolf Gebhardt, Stefan Berger, Flavonoid Binding to a Multi-Drug-Resistance Transporter Protein: an STD-NMR Study, 2004.
[39] S.R. Brozell, S. Mukherjee, T.E. Balius, D.R. Roe, D.A. Case, R.C. Rizzo, Evaluation of DOCK 6 as a pose generation and database enrichment tool, J. Comput. Aided Mol. Des. 26 (2012) 749–773.
[40] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854.
[41] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for

molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[42] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Eng. Des. Sel. 8 (1995) 127–134.
[43] R. Kumari, R. Kumar, A. Lynn, A. Lynn, *g_mmpbsa* —a GROMACS tool for high-throughput MM-PBSA calculations, J. Chem. Inf. Model. 54 (2014) 1951–1962.
[44] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[45] D. Leckband, J. Israelachvili, Intermolecular forces in biology, Q. Rev. Biophys. 34 (2001) 105–267. http://www.ncbi.nlm.nih.gov/pubmed/11771120. (Accessed 8 October 2019).
[46] B. Gay, S. Suarez, G. Caravatti, P. Furet, T. Meyer, J. Schoepfer, Selective GRB2 SH2 inhibitors as anti-Ras therapy, Int. J. Cancer 83 (1999) 235–241.
[47] F.-D.T. Lung, J.-Y. Tsai, Grb2 SH2 domain-binding peptide analogs as potential anticancer agents, Biopolymers 71 (2003) 132–140.
[48] K. Machida, B.J. Mayer, The SH2 domain: versatile signaling module and pharmaceutical target, Biochim. Biophys. Acta Protein Proteonomics 1747 (2005) 1–25.
[49] W.C. Shakespeare, SH2 domain inhibition: a problem solved? Curr. Opin. Chem. Biol. 5 (2001) 409–415.
[50] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), Proteomics Protoc. Handb., Humana Press, Totowa, NJ, 2005, pp. 571–607.
[51] F.P. Benzati, I.P. Caruso, T.S.P. Teixeira, F.R. Moraes, F.A. Melo, L.C. Di Stasi, M.L. Cornelio, F.P. Souza, M.A. Fossey, Investigation of the interaction between Coumarins and its derivatives with human serum albumin: STD-NMR, fluorescence and docking molecular studies, Intermt. J. Sci. 2 (2016) 1–11.
[52] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The protein data bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[53] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, I.D. Kuntz, DOCK 6: combining techniques to model RNA–small molecule complexes, RNA 15 (2009) 1219–1230.
[54] T.J.A. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, J. Comput. Aided Mol. Des. 15 (2001) 411–428.
[55] V. Zoete, M.A. Cuendet, A. Grosdidier, O. Michielin, SwissParam: a fast force field generation tool for small organic molecules, J. Comput. Chem. 32 (2011) 2359–2368.
[56] J.E. Ladbury, M. Doyle, Biocalorimetry II: Applications of Calorimetry in the Biological Sciences, John Wiley & Sons, 2004. https://www.wiley.com/en-br/Biocalorimetry+2:+Applications+of+Calorimetry+in+the+Biological+Sciences-p-9780470849682. (Accessed 8 October 2019).

Journal of Molecular Structure 1234 (2021) 130164

Contents lists available at ScienceDirect

Journal of Molecular Structure

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

Thermodynamic profile and molecular modeling of the interaction between Grb2 dimer and flavonoids Rutin and Morin

Karoline Sanches[a,b,1], Raphael V.R. Dias[a,1], Paulo H. da Silva[a,b], Icaro P. Caruso[a,b,c], Marcelo A. Fossey[a,b], Fátima P. de Souza[a,b], Leandro C. de Oliveira[a], Fernando A. Melo[a,b,*]

[a] Department of Physics — Institute of Biosciences, Humanities and Exact Sciences (IBILCE) São Paulo State University "Júlio de Mesquita Filho" (UNESP), 15054-000 São José do Rio Preto, SP, Brazil
[b] Multiuser Center for Biomolecular Innovation (CMIB), Institute of Biosciences, Humanities and Exact Sciences (IBILCE), São Paulo State University "Júlio de Mesquita Filho" (UNESP), 15054-000 São José do Rio Preto, SP, Brazil
[c] Institute of Medical Biochemistry Leopoldo de Meis (IBqM) and National Center for Structural Biology and Bioimaging (CENABIO), Federal University of Rio de Janeiro (UFRJ), 21941-590 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 11 September 2020
Revised 18 February 2021
Accepted 19 February 2021
Available online 24 February 2021

Keywords:
Interaction
Grb2
Flavonoids
Thermodynamics
Molecular modeling

ABSTRACT

The adaptor protein growth factor-bound protein 2 (Grb2) is an important regulator of the fibroblast growth factor receptor 2 (FGFR2) before extracellular stimuli. It is known to form complexes that end up in the mitogen-activated protein kinase (MAPK) pathway activation, which is involved in proliferation and oncogenic signal transduction. Grb2 is a versatile protein performing functions other than adaptor protein, making it a relevant target to verify its interaction with flavonoids such as Rutin and Morin. These small polyphenols molecules are easy to be found in the nature and its anti-tumor properties are well-known. By using fluorescence spectroscopy, the thermodynamic profile of the interaction between those molecules and Grb2 showed entropically driven interactions, where hydrophobic effects take place as the main interaction potential. The dissociation constants found were K_d ~10^{-6} M for Morin and K_d ~10^{-5} M for Rutin. The molar ratio protein/ligand is 1:1 for both assays. Furthermore, nuclear magnetic resonance has provided important information about the protein-ligand interaction epitopes, which has been used as a guide for molecular docking and molecular dynamics simulations. The combination of the obtained results shows the SH2 domain as the most probable interaction place on Grb2 dimer for Rutin and Morin binding. Sh2 is a well-known domain responsible for pY (phosphotyrosine) recognition upon protein partners and an important protein module for testing SH2 domain inhibitors.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Cell metabolism is mediated by signaling transduction, which is dependent upon reversible phosphorylation of the amino acid side chains [19]. The Fibroblast growth factor receptor 2 (FGFR2) is a crucial receptor tyrosine kinase (RTK) that regulates cell metabolism, gene expression, cell growth, division, and differentiation [1,12]. Lin et al. [23] showed that the regulation of FGFR2 is made by the adaptor protein Grb2 (Growth factor-bound protein 2) that prevents FGFR2 from reaching full activation before extracellular stimulation through growth factors binding [23]. Grb2 binds as a dimer to the last fifteen C-terminal proline-rich amino acids motif (PXXP; X stands for any other amino acid residue) found in FGFR2 through the Grb2 C-terminal SH3 domain. A heterotetrameric formation occurs favoring the transphosphorylation of tyrosine residues along the 58 amino acids chain of the C-terminal region of FGFR2 [23,28]. When a growth factor binds to the FGFR2 extracellular region, it induces conformational changes in the cytoplasmic region of FGFR2 and consequently phosphorylation of Grb2. It causes Grb2 to dissociate from FGFR2, which, in turn, begins to recruit proteins from the cytosol to initiate early signaling complexes. This event indirectly activates the mitogen-activated protein kinase (MAPK) signaling pathway, which can lead to disorderly cell growth related to a variety of human cancers and developmental defects [5,10,23,41,43]. Another study has reported that Grb2 has a monomer/dimer equilibrium, which is pivotal for the MAPK activation [2]. Grb2 monomer binds to the son of sevenless protein (SOS) to regulate the MAPK activity, while its dimeric form is inhibitory on this process. Grb2 has a pivotal role in regulating cell signaling-related events that will develop cancer and other human

* Corresponding author.
E-mail address: fernando.melo@unesp.br (F.A. Melo).
[1] The authors contributed equally.

https://doi.org/10.1016/j.molstruc.2021.130164
0022-2860/© 2021 Elsevier B.V. All rights reserved.

developmental disorders and can be a new target to test molecules that present antitumor properties in cell assays.

Over the last few years, there has been a growing interest in natural compounds with pharmacologic potential and low toxicity to interact with target proteins in disordered signaling pathways [4,9]. The flavonoids family is a large group of secondary metabolites of the polyphenol class, naturally present in plants. These classes of compounds have been widely studied due to their benefits to human health, such as anticarcinogenic, antioxidant, anti-inflammatory, and antiviral properties [9]. Among those compounds, we can highlight Rutin (*Quercetin-3-O-α-L-Rhamnopyranosyl-(1→6)-β-D-Glucopyranoside*) and Morin (*2′,3,4′,5,7-pentahydroxyfavone*). The first one is a glycosylated flavonoid found in typical plants, such as buckwheat, passion flower, apple, and tea, and known to induces apoptosis via the mitochondrial pathway in human colon cancer cells, decreasing cell viability in a concentration-dependent manner [8,24,31]. It has been shown that the early treatment with Rutin in lung tissues inhibits histopathological changes [24]. Also, it was been reported the anticarcinogenic activity of Rutin in cancer cells induced by the MAPK pathway [11,25,29]. The second one is a flavonoid isolated from various plant species, such as watercress, osage orange, leaves of guava, and olive leaves [17,42]. Studies have indicated that Morin inhibits cell proliferation in several types of carcinogenic cells by inhibiting the cell cycle in tumors such as oral carcinoma and human hepatocytes [18,35]. Despites of all these studies, showing the anti-tumor properties of Rutin and Morin and the pivotal role of Grb2 on the cell proliferation mediating process, a molecular level characterization of the interaction among the flavonoids and Gb2 are missing. In this study, we have used a combined experimental and theoretical biophysical methods, such as fluorescence spectroscopy, STD-NMR, molecular dynamics, and molecular docking to characterize the interaction of Grb2 dimer and the molecules. Our results have shown that both Morin and Rutin compete for the same binding site on the Grb2-SH2 domain which is a well-known and versatile signaling module for drug targeting [26].

2. Results and discussion

2.1. Grb2 dimer and flavonoids interaction analysis through fluorescence

The fluorescence quenching intensity occurs by a variety of events, including excited state reactions, molecular rearrangements, energy transfer, ground state complexation, and collisional quenching [21]. The analysis of Grb2 fluorescence quenching while titrating increasing amounts of ligands in different temperatures (Fig. 1) have shown that the higher concentration of flavonoid into the protein microenvironment, lower is the Grb2 fluorescence intensity. For Morin and dimeric Grb2, the fluorescence intensity decreases around 55%, ranging from 110 to 60 a.u. (Fig. 1A on the top right) while Rutin makes the fluorescence intensity decrease around 70% on the same experimental arrangement, ranging from 160 to 110 a.u. (Fig. 2A on the top right). Grb2 has five tryptophans residues which are spread as following: two in the SH2 domain, two in the C-SH3 domain, and one in the N-SH3 domain. Since there is no 100% quenching of the protein Trp, we can suppose that the ligands are binding to Grb2 dimer in a particular place where one or more Trp-residue is being perturbed by the presence of the flavonoid.

The stoichiometry n and dissociation constant K_d where determined through the double-log model (Eq. (1)) by following the analysis of the fluorescence quenching during flavonoid titrations [7,11].

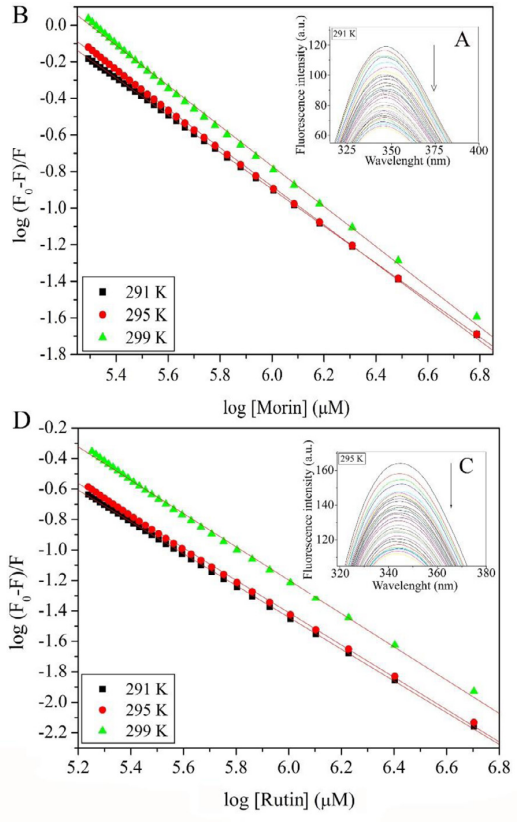


Fig. 1. Fluorescence emission spectra of Grb2 protein in the presence of flavonoids and double-log plot model of the interaction. (A) and (C) show the intrinsic fluorescence quenching of Grb2 tryptophan residues in the presence of increasing amounts of flavonoids Morin and Rutin respectively. (B) and (C) show the double-log graph. The linearity behavior for the experiments performed in three different temperatures (291, 295 and 299 K) revels that the interaction of both flavonoids with Grb2 agree with the model, where one ligand interacts with one Grb2 protomer. The dissociating constants was calculated to be K_d ~10^{-6} M for Morin and 10^{-5} M for Rutin.

$$log\left[\frac{(F_0 - F)}{F}\right] = nlog{}^{1}/_{K_d} + nlog[L] \quad (1)$$

In the above equation, F_0 and F are the tryptophan intrinsic fluorescence before and after titrating flavonoid in the system respectively while [L] is the concentration of flavonoid after each titration.

The double-log model was chosen to analyze the experiments due the linear dependance of fluorescence quenching with flavonoid concentration during titrations. From the results shown above (Fig. 1B and D) the stoichiometry n could be determined from graph slope and the K_d could be determined in the graph where the line intercepts the Y-axis. These experiments were performed at three different temperatures and the results can be found in Table 1. For the two sets of experiments, we can see that the dissociation constant is temperature dependent as this parameter decrease while temperature increase on both cases. This behavior shows that the temperature is an important physicochemical parameter to make the interaction more favorable an to stabilize the protein-ligand complex.

The formation and stabilization of protein-ligand interactions are driven by van der Waals forces, electrostatic forces, hydrophobic effects, etc. [7]. Since we have obtained the dissociation con-

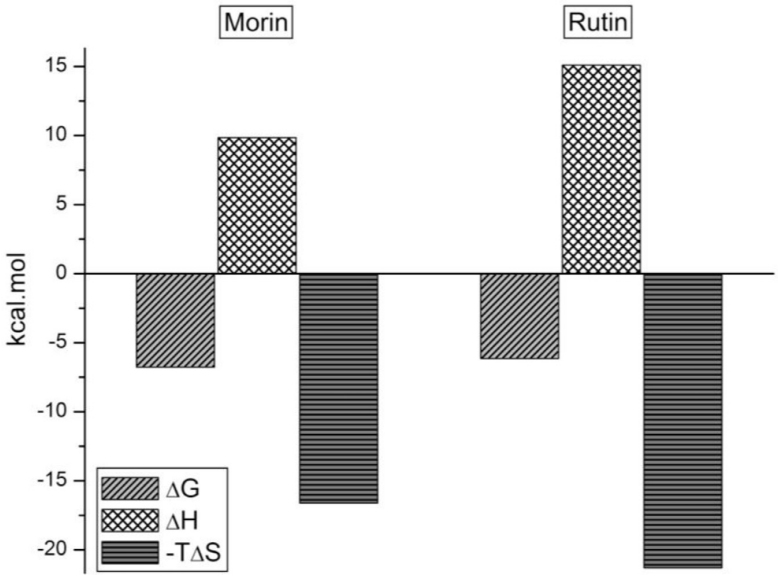


Fig. 2. Thermodynamic profile of the interaction between Grb2 and Morin and Rutin flavonoids. Both interactions are entropically driven.

Table 1
Numbers of ligand per protein *n*, the dissociation constant K_d and the thermodynamic parameters.

Flavonoid	T (K)	n	K_d (μM)	ΔG (kcal.mol^{-1})	ΔH (kcal.mol^{-1})	TΔS (kcal.mol^{-1})
Morin	291	1.01	7.9	−6.75	9.85	16.6
	295	1.05	7.0	−6.95		16.8
	299	1.09	5.0	−7.25		17.1
Rutin	291	1.04	23	−6.2	15.1	21.3
	295	1.06	20.3	−6.4		21.5
	299	1.09	11.4	−6.8		21.9

stant K_d of the complex formed by Grb2 dimer and the aforementioned flavonoids in three different temperatures, we have used the van't Hoff analysis (Eq. (2)) to infer the main forces are taking place during the interaction.

$$ln^1/_{K_d} = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \quad (2)$$

In the above equation, R is the ideal gas constant (8.31 J.mol^{-1}.K^{-1}), K_d is the dissociation constant at the correspondent temperature (T) in degrees Kelvin. ΔH and ΔS are the enthalpy and entropy changes, respectively. The corresponding Gibbs free energy (ΔG) contribution were calculated from the Eq. (3). All these parameters can be found in Table 1 for the two sets of experiments.

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

The thermodynamic profile of the interaction between Grb2 dimer and flavonoids are shown (Fig. 2) as an average of the energetic contributions of Gibbs free energy (ΔG), enthalpy (ΔH), and temperature times entropy (TΔS).

According to the above thermodynamic profile we can see that the free energy is negative ($\Delta G < 0$), which means that the interaction of Grb2 dimer with Morin or Rutin is a spontaneous process on both cases. It is already known that negative values for ΔH and TΔS imply in van der Waals forces and hydrogen bonding, while positive values are found when hydrophobic effects are dominant [37]. In the case of the Grb2 interaction with Morin and Rutin, our results showed $\Delta H > 0$ and $T\Delta S > 0$. Besides, the thermodynamic profile of Fig. 2 is a classic picture which the major contribution for the complex formation is due to hydrophobic events despite other contributions as van der Waals interaction and hydrogen bonding can happen in small favorable amounts [16]. This behavior agrees with the hydrophobic nature of the ligands used in this study. The higher and negative contribution of the entropic term for the Gibbs free energy can be understood as displacement/rearrangement from the site specifically bound water molecules while hydrophobic forces taking place during the interaction. On the other hand, the positive enthalpic contribution is related with changes in solvent accessible surface area upon ligand binding followed by a local protein conformational change, that is an unfavorable process. This suggests that ligand induce a local conformational change while burying itself in the binding site [20,33,45].

2.2. Saturation transfer difference based nuclear magnetic resonance

STD-NMR spectroscopy experiments was used to identify the epitopes of interaction between Grb2 dimer and flavonoids Morin and Rutin and to verify the site competition between both ligands. The hydrogen of the polyphenolic compounds was identified following the assignment available in Spectral *Database* for Organic Compounds, *SDBS* (https://sdbs.db.aist.go.jp). The difference between signals (Fig. 3), reflects the hydrogen atoms that received the saturation transfer during protein-ligand interaction [46]. The epitopes of the interaction were calculated following the Eq. (4):

$$I = \frac{I_{STD}}{I_0} = \frac{I_0 - I_{SAT}}{I_0} \quad (4)$$

where I_0 is the intensity of the reference spectrum (off-resonance), I_{SAT} is the intensity saturation spectrum and I_{STD} is the intensity of

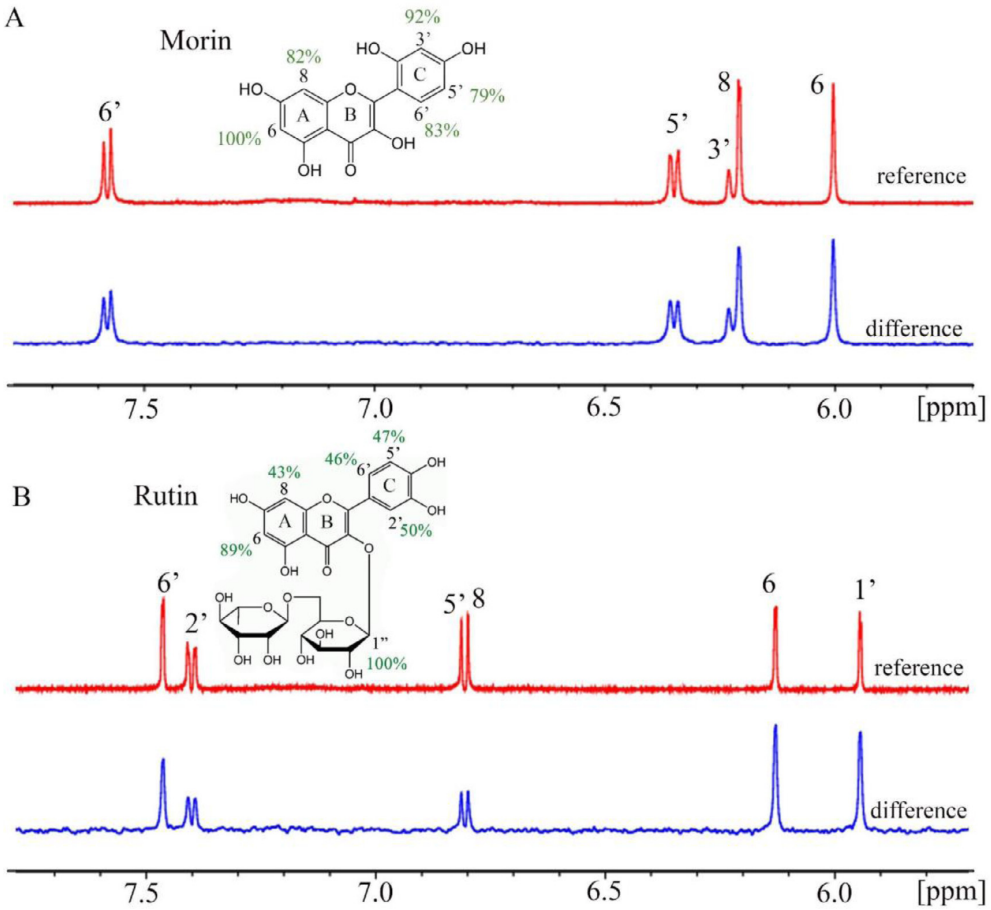


Fig. 3. STD spectrum Grb2-Rutin complex where the off-resonance (red) is the experiment which there is no perturbation of any signal of the protein or ligand, and the blue spectra refer the saturation that the ligand received from the protein during the interaction. In (A) it is seen the signals from the Grb2-Morin interaction and in (B) the signals from the Grb2-Rutin interaction. The calculated percentage in relation of the saturation transfer of each ligand is found in green at the flavonoid structures above. Those values are normalized.

Table 2
Epitopes of interaction between Grb2 and the flavonoids Rutin and Morin with the hydrogen and its respective position δ.

	Individual	Competition			
Flavonoid	Hydrogen	δ (ppm)	I	I	Epitope difference (%)
Rutin	1''	5.94	0.10626	0.09609	36
	6	6.12	0.09438	0.08771	30
	2′	7.46	0.05339	0.04949	17
	5′	6.81	0.04952	0.04625	16
	6′	6.40	0.04947	0.05267	11
	8	6.79	0.04545	0.04174	15
Morin	6	6.00	0.16107	0.14078	5
	3′	6.23	0.14859	0.13491	1
	6′	7.58	0.13378	0.14818	17
	8	6.21	0.13219	0.12864	5
	5′	6.35	0.12761	0.12756	7

the difference spectrum [15]. The epitope values for the interaction between Grb2 and the flavonoids can be found in Table 2.

The hydrogen 1″ of Rutin and hydrogen 6 of Morin are the atoms that receive the higher amount of magnetization from Grb2 while hydrogen 8 of Rutin and 5′ receive the lower amount of magnetization from Grb2 when these interactions are analyzed individually. So, these results show the way of how the ligands orient themselves to the Grb2 binding site during the interaction. Besides there are no observed signals coming from the sugar hydrogens which suggest a steric hindrance, where the disaccharide would not fit in the hydrophobic protein pocket [15].

Also, it was verified the competition between the two polyphenols and calculated how the ligand epitope affects each other during the interaction. The effective value of the perturbation of the

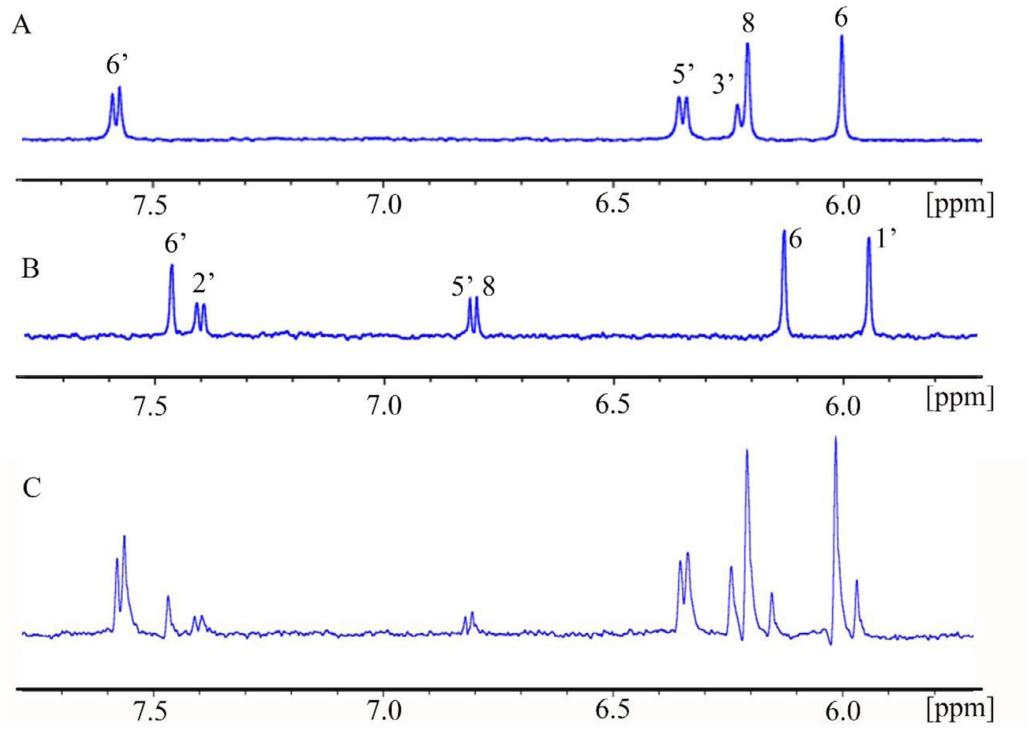


Fig. 4. STD spectrum of (A) Rutin, (B) Morin, and (C) the competition between both flavonoids for the binding site in Grb2.

competitor by the following Eq. (5):

$$I_{af} = \frac{\Sigma E_t}{\Sigma I} 100\% \quad (5)$$

ΣEt is the sum of the product of each epitope and the percentage of the competition to the individual epitope and ΣI is a ponderation factor. And by multiplying the division for 100% it was obtained: This is a simple statistical method to understand how one ligand is perturbed by the presence of the other and to get the percentage of this perturbation.

In Fig. 4 it can be observed the difference spectra of Morin and Rutin individually, and the competition spectra between both molecules.

To better understand the above experiment the values of each hydrogen epitope of Rutin and Morin (A and B) interacting individually with Grb2, and the epitope of a mixture of Rutin and Morin (molar ratio 1:1) competing for the same Grb2 binding site were calculated and they are presented in the Table 2.

To demonstrate the higher affinity of the Morin, we normalize the Grb2-Morin individual complex to 100%, verifying that in the competition it had a reduction to 93.2%, meaning that Rutin affects 6.8% the interaction between Grb2 and Morin. On the other hand, considering Grb2-Rutin individual complex 100%, in the competition experiment it was reduced to 75%, showing that 25% of the Grb2-Rutin interaction is affected by Morin. Then, we can observe that Morin is less affected by Rutin, suggesting that the Grb2-Morin complex affinity (K_d) is higher when compared to Grb2-Rutin. This result corroborates with the fluorescence experiments.

2.3. Computational evaluation of interactions

A detailed search of possible pockets for the interaction between the Grb2 and flavonoids (denominated cluster by the UCSF Dock terminology) indicated possible regions with physical requirements (volume and charge) to receive the flavonoids. Among these clusters, there was one located in the SH2 domain with potential to be the most suitable when compared to the experimental results, being its characteristics: predominantly hydrophobic, located in an accessible protein region, containing a tryptophan residue within the pocket (Trp 121) and already highlighted in the literature as a receptive domain to small molecules [30,36,38–40] (Fig. 5A).

Since a candidate cavity was defined, molecular docking was performed for the Morin flavonoid. After 1000 runs, the molecular docking showed a predominance [around 93%] of a conformation where Morin rings A and B, which form the most rigid structure fit into the hydrophobic pocket of the SH2 domain, whereas the ring C which is a more flexible ring does not completely fit into the cavity (Fig. 5B and C). 50 ns of molecular dynamics was performed to ensure that the found conformation and interaction remains stable when the SH2 domain and the ligand are flexible. The result showed a permanence of the ligand throughout molecular dynamics, with small adjustments of the side chains and small movements of the C ring of the flavonoid, without interfering in the interaction frequency results.

The final conformation taken from the molecular dynamics as being the one with the lowest energy (van der Waals potential) follows the results obtained by STD-NMR (Fig. 3), with the carbon-bound hydrogen corresponding to the epitope with intensity of 100% being the most buried in the cavity. The interactions between protein and ligand show a predominance of hydrophobic interactions, with a possible hydrogen bond formed by Trp121 (Fig. 5D), also corroborating the results of fluorescence.

All processes previously described for the Morin flavonoid were repeated for Rutin and, although they presented the same basic structure, they differ only by the sugar. The molecular docking of

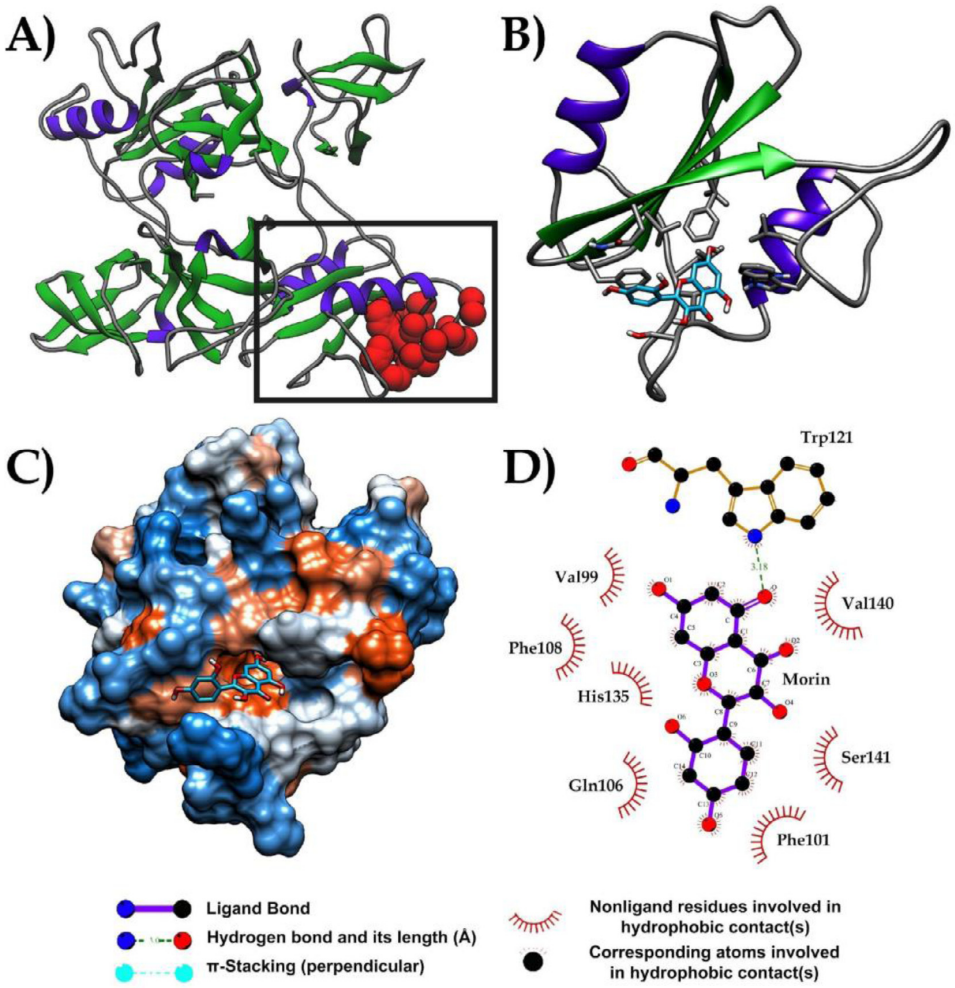


Fig. 5. (A) Grb2 in its dimeric form, highlighting the pocket in domain SH2 with the red spheres of rays 1.4 Å representing the possible anchorage sites for the ligand. (B and C) Best-ranked structure of the SH2-Grb2 complex with the Morin ligand. This structure shows a fitting of the rigid Morin rings fitting into the hydrophobic cavity of the site. The representations were generated in cartoon for visualization of the contacts generated and in surface to facilitate the visualization of the hydrophobic pocket. (D) Possible interactions of the amino acid residues of Grb2 with Morin. These contacts were calculated according to the characteristics of each residue and using a distance range of up to 4 Å. The images and analyzes were performed using the software Chimera [32] and LigPlot [44].

Rutin presented differences when compared to Morin. The Rutin showed the predominance of a conformation in which its rings A and B do not fully fit into the hydrophobic pocket (Fig. 6A and B). This result probably is due to steric constraints caused by the sugar branching in its main rings. These analyzes show that because Rutin is larger than Morin, the constraints due to residues in its structure make its interaction more superficial than the Grb2-Morin interactions, differing on average by approximately 1 Å between the A-rings of both flavonoids compared to tryptophan 121 in the final structure. This information explains and validates the results of competitions measured by STD-NMR between the flavonoids: the fact that Rutin is larger than Morin implies a disadvantage in fitting, making the Grb2-Morin interaction more likely in the competition. Besides, the presence of a tryptophan residue in the cavity, hydrophobic predominance and fitting orientation are results that corroborate with fluorescence and NMR-STD experiments (Fig. 6C and D).

3. Conclusion

Grb2 is an important protein ubiquitously expressed inside the cell. Besides being an adaptor protein, Grb2 is also a global regulator of FGFR2 kinase activity before extracellular stimulus [23]. Also, the equilibrium between monomeric and dimeric states of Grb2 works like a switch upon the regulation of the MAPK pathway. The monomeric Grb2 up-regulates MAPK pathway while dimeric Grb2 would do the opposite [2]. So, the ability of this protein to mediating cell signaling transduction through its SH2 and SH3 domains makes Grb2 an important protein target to test the interaction with several molecules that present antitumor properties. Our results have shown that Rutin and Morin competes for the same hydrophobic binding site located in the SH2 domain. It is worth mentioning that other domains of Grb2 are capable of interacting to these kinds of molecules (data not shown), but when Grb2 dimer takes place in the system only SH2 domain would

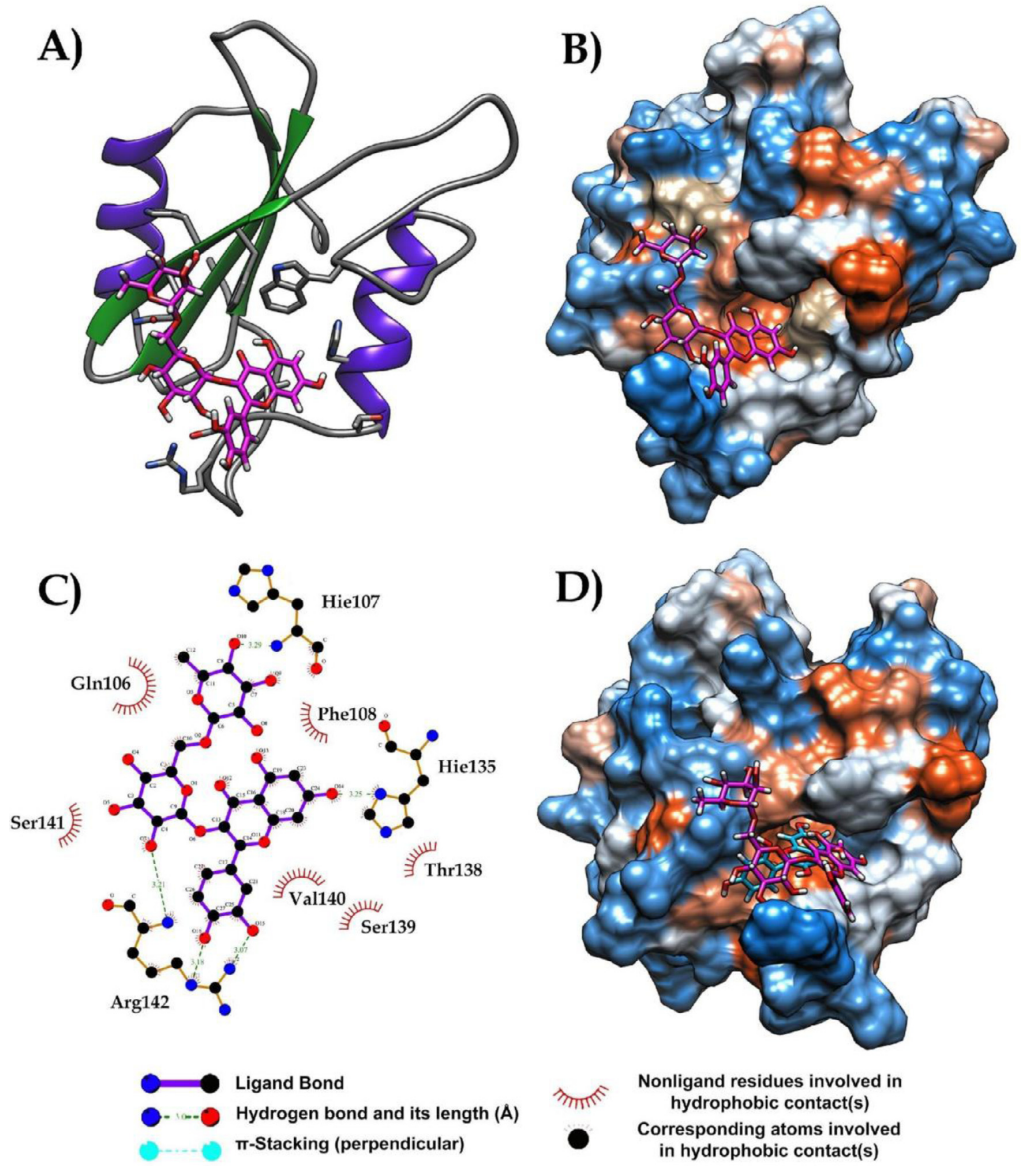


Fig. 6. (A) and (B) Grb2-Rutin complex obtained by molecular docking. This result shows an anchorage in the hydrophobic site, and its rings could not fully fit into the cavity due to steric constraints caused by the sugar. (C) Possible interactions between amino acid residues and Rutin, calculated using as parameters the residues loading and their distances. (D) Overlapping of the Morin and Rutin ligands, here it is possible to visualize the differences of the fittings, where Morin is most fitted to the hydrophobic cavity.

be available to receive the molecules according to the results discussed above. This binding site was already shown to interact with coumarin 1,2-benzopyrone [40], a phenolic molecule that has been used in many applications such as: antibiotic, anti-inflammatory, anticoagulant, vasodilator, anti-tumor, among others [13]. This is just to show that Grb2 and SH2 domain is a worth valued macromolecule to test novel pharmacological compounds against cancer. While Morin fits into the hydrophobic cavity found in the SH2 domain Rutin interacts to the SH2 domain more superficially due steric constraints. These results explain why Morin takes advantage over Rutin in the competition experiments. Our findings could level up Grb2 as an important target to test the interaction with polyphenols molecules already known to have anticancer properties. Also, a study about how these phenolic molecules disturb the Grb2 mediating cell signaling are still missing. Dimeric Grb2 can be disrupted either by repulsive electrical forces when Y160 on C-SH3 domain is phosphorylated or through steric clash effects created by interactions of small phosphopeptides interacting in the SH2 domain [2]. Despites still there is no evidence that Morin and Rutin can promote Grb2 dimer dissociation during interaction, we still can think in a situation where those molecules can possibly work somehow as a SH2 domain inhibitor and block the link for Ras/MAPK signaling transduction to prevent aberrant cell proliferation as well as cancer outcome.

4. Materials and methods

4.1. Expression and purification of Grb2 protein

The 6x histidine-tagged Grb2 was expressed in *E. coli* BL21 (DE3) Gold. The cell culture has grown in LB media containing 50 μg/mL of kanamycin and 50 μg/mL of tetracycline by shaking at 120 RPM and 37 °C until OD_{600} = 0.6. The temperature was lowered to 20 °C and the protein expression was induced by the addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The culture was left to shaking at 120 RPM for 12 h at 20 °C. Then, the cells were harvested by centrifugation at 4000xg for 20 min at 20 °C and resuspended in buffer A (50 mM Tris–HCl, 300 mM NaCl, 1 mM β-Mercaptoethanol (βME), 0.25 mg/mL of lysozyme, 5 mM imidazole, protease inhibitor (P2714–1BTL-Sigma Aldrich®), pH 8.0). After lysis by sonication, the cells were centrifuged at 34.957xg and 10 °C for 50 min. The supernatant was filtered in 0.22 µm Minisart® filter and loaded into metal affinity column IMAC HiTrap HP (GE Life Sciences) equilibrated with buffer A. After washing the column with five volumes of buffer A, Grb2 was eluted with buffer B (50 mM Tris–HCl, 100 mM NaCl, 1 mM βME, 200 mM imidazole, pH 8.0). The sample was concentrated to 1.0 mL and loaded into a Superdex 200 10/30 GL column (GE Life Science) equilibrated in buffer C (20 mM NaH_2PO_4, 50 mM NaCl, 1 mM βMe, pH8.0). Analysis of the purity of the protein was verified by 15% SDS-PAGE.

4.2. Grb2 and flavonoids stock solutions

It was used Thermo Scientific BIOMATE UV-Visible (Thermo Fisher Scientific) spectrophotometer equipped with a quartz cell of 1.0 cm path length, scanning speed of 600 nm/min, 1.0 nm of interval and spectral bandwidth of 2.0 nm to determine the Grb2 and flavonoids stock solution. The spectra were performed at room temperature. The UV–Vis absorption spectrum was recorded from 200 to 500 nm. The concentration calculation was done according to the Beer-Lambert equation using the molar extinction coefficient at 280 nm of $\varepsilon_{280} = 38.055\ M^{-1}cm^{-1}$ for Grb2, $\varepsilon_{363} = 14,500\ M^{-1}cm^{-1}$ for Rutin and $\varepsilon_{271} = 16,949\ M^{-1}cm^{-1}$ for Morin [3,11]. Both flavonoids were diluted in absolute ethanol.

4.3. Fluorescence steady state

The fluorescence measurements were performed on the Cary Eclipse Varian® Spectrofluorometer, equipped with Peltier Single Cell Holder System, in quartz cuvettes of 1 cm optical path, at temperatures of 291, 295 and 299 K, to perform the thermodynamic characterization of the interactions, with an opening of the emission and excitation slits set at 0.5 mm. The tryptophans of the Grb2 protein were excited at 290 nm with the emission range collected from 300 to 500 nm. Fluorescence quenching was evaluated by filling a quartz cuvette with 2 μM Grb2 in 2000 μL. Then, 2 μL of flavonoid from a stock solution was titrated in each experiment. The total sample volume changed 3% while ethanol concentration into the sample was less than 0.25% v/v. Flavonoids concentration changed during experiments in a range of 0 to 5.9 μM with increments of 0.2 μM per titration.

4.4. Saturation transfer difference (STD)

NMR spectra were collected on a Bruker Avance III 600.13 MHz spectrometer (Bruker, Germany) equipped with a 5 mm triple resonance cryoprobe, with the field gradient pulsed along the Z-axis. All data were processed and analyzed with Bruker Topspin version 3.2. For evaluating the interaction, it was used the Saturation Transfer Difference (STD) by NMR using pulse sequence STD-DIFFESGP. The protein selective saturation is performed and compared to a reference spectrum. Grb2 solution at 5 μM, pH 8.0 in phosphate buffer was used to determine saturation conditions. Protein saturation at −1.0 ppm (on-resonance) was selected, keeping the off-resonance at 20 ppm. The saturation time of Grb2 was 5 s. Rutin and Morin were added at a concentration of 200 μM. The equilibrium time was 6 s for the Grb2 flavonoids complex. A total of 512 scans was collected for and the experiment was carried out at 20 °C.

4.5. Molecular docking and molecular dynamics simulations

Molecular docking simulations of the Grb2-Coumarin complex were performed employing the tridimensional information deposited in the Protein Data Bank [6] (PDBid.: 1GRI). Morin and Rutin (2′,3,4′,5,7-pentahydroxyfavone and Quercetin-3-O-α-L-Rhamnopyranosyl-(1→6)-β-D-Glucopyranoside respectively) structures were obtained from PubChem database (https://pubchem.ncbi.nlm.nih.gov/) with CID. 5280805 and 5281670 and both were prepared using UCSF Chimera Tools [32]. The grid box used was built using the distance of 8 Å from the center of each cluster in the x, y, and z directions. This larger box is divided into bins of 0.2 Å and the distance tolerance for matching ligand atoms to the receptor set in 0.75 Å.

The first step of molecular docking was performed using the single grid energy (SGE) score function composed by the electrostatic and van der Waals interaction terms, this calculation is performed keeping the protein rigid and the flexible ligand using the “anchor and grow” algorithm [14]. Each docking run tests 500 possible ligand orientations, selecting only the lowest energy. This process is repeated 1000 times, generating a total of 1000 conformations allowing a statistical analysis of the generated conformations. Completely, the AMBER Score function [22] was employed for a rescoring of the all conformations and final energy calculation. This second method is important since it allows degrees of flexibility in the pocket and ligand.

The Molecular Dynamics simulations were carried out using the GROMACS package version 4.5.5 [34]. The force field CHARMM27 [27,34] with the standard set up for the protein residues and TIP3P water model was employed. Rutin and Morin ligands were parameterized using Swissparam webserver from the Swiss Institute of Bioinformatics [47]. Energy minimization was done using 50,000 steepest-descent and 5000 conjugate gradient steps, both without position restraints. To the system equilibration, the first step of 1 ns with restrictions on protein and ligand positions was performed, followed by another 1 ns without restrictions. The MD simulation was done for 20 ns with an integration step of 2 fs at 295 K, 1 atm, and salt concentration of 0.15 M. The trajectory positions and energies were recorded each 100 ps. The simulation box was built with 10 Å of distance from the protein surface (x,y, z-axis) with Periodic Boundary Conditions (PBC). LYNCS method was used to constrain all hydrogen bonds. Parrinello-Rahman barostat and Berendsen thermostat were used for pressure and temperature control respectively. Since molecular simulations and docking calculations employs different energy expressions, the conformation with the lowest energy from simulations were submitted to a complete molecular docking procedure (SGE and ASBE), providing comparable energies.

Author contributions

P.H.S. and K.S. performed the experiments. R.V.R.D and L.C.O. planned and carried out the computational simulations. F.A.M., I.P.C., M.A.F., F.P.S. made the project idea and supervised the experimental work. F.A.M., K.S. and R.V.R.D. wrote the manuscript. All

authors contributed with the analysis, discussion and critical feedback.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.as potential competing interests:

Acknowledgment

This work was supported by FAPESP grant n°. 2014/17630–0, 2016/08753–6, 2019/24974–0 and CNPq grant n°. 442951/2014–0 and 442352/2014–0. The simulations were performed at the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) and CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em São Paulo), project UNICAMP/FINEP-MCT. Scholarships were supported by CAPES and CNPq and FAPESP.

We thank the support of Prof. Dr. John E. Ladbury (University of Leeds, United Kingdom) who provides the Grb2 plasmid vector. Dr. Fábio R. de Moraes (Multiuser Center for Biomolecular Innovation, São Paulo State University) to support collect the NMR data and analyses. Prof. Dr. Márcio F. Colombo (São Paulo State University) for the Laboratory of Spectroscopy.

References

[1] I. Ahmad, T. Iwata, H.Y. Leung, Mechanisms of FGFR-mediated carcinogenesis, Biochim. Biophys. Acta - Mol. Cell Res. 1823 (2012) 850–860.
[2] Z. Ahmed, Z. Timsah, K.M. Suen, N.P. Cook, G.R. Lee, C.-.C. Lin, M. Gagea, A.A. Marti, J.E. Ladbury, Grb2 monomer–dimer equilibrium determines normal versus oncogenic function, Nat. Commun. 6 (2015) 7354.
[3] A. Aitken, M.P. Learmonth, Protein determination by UV absorption, in: The Protein Protocols Handbook, Springer Protocols Handbooks. Humana Press, Totowa, NJ, 2009, pp. 3–6.
[4] S. Attoub, A.H. Hassan, B. Vanhoecke, R. Iratni, T. Takahashi, A.-.M. Gaben, M. Bracke, S. Awad, A. John, H.A. Kamalboor, M.A. Al Sultan, K. Arafat, C. Gespach, G. Petroianu, Inhibition of cell survival, invasion, tumor growth and histone deacetylase activity by the dietary flavonoid luteolin in human epithelioid cancer cells, Eur. J. Pharmacol. 651 (2011) 18–25.
[5] I.S. Babina, N.C. Turner, Advances and challenges in targeting FGFR signalling in cancer, Nat. Rev. Cancer 17 (2017) 318–332.
[6] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The protein data bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[7] S. Bi, L. Ding, Y. Tian, D. Song, X. Zhou, X. Liu, H. Zhang, Investigation of the interaction between flavonoids and human serum albumin, J. Mol. Struct. 703 (2004) 37–45.
[8] S.L. Bondarev, V.N. Knyukshto, Fluorescence and phosphorescence of rutin, J. Lumin. 142 (2013) 236–240.
[9] L.H. Cazarolli, L. Zanatta, E.H. Alberton, M.S.R.B. Figueiredo, P. Folador, R.G. Damazio, M.G. Pizzolatti, F.R.M.B. Silva, Flavonoids: prospective drug candidates, Mini. Rev. Med. Chem. 8 (2008) 1429–1440.
[10] A. Desai, A.A. Adjei, FGFR signaling as a target for lung cancer therapy, J. Thorac. Oncol. 11 (2016) 9–20.
[11] F. Ding, G. Zhao, J. Huang, Y. Sun, L. Zhang, Fluorescence spectroscopic investigation of the interaction between chloramphenicol and lysozyme, Eur. J. Med. Chem. 44 (2009) 4083–4089.
[12] S.E. Egan, B.W. Giddings, M.W. Brooks, L. Buday, A.M. Sizeland, R.A. Weinberg, Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation, Nature 363 (1993) 45–51.
[13] S. Emami, S. Dadashpour, Current developments of coumarin-based anti–cancer agents in medicinal chemistry, Eur. J. Med. Chem. 102 (2015) 611–630.
[14] T.J.A. Ewing, S. Makino, A.G. Skillman, I.D. Kuntz, DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, J. Comput. Aided. Mol. Des. 15 (2001) 411–428.
[15] D. Haiyun, C. Jianbin, Z. Guomei, S. Shaomin, P. Jinhao, Preparation and spectral investigation on inclusion complex of beta-cyclodextrin with rutin, Spectrochim. Acta A Mol. Biomol. Spectrosc. 59 (2003) 3421–3429.
[16] I. Haq, Thermodynamics of drug-DNA interactions, Arch. Biochem. Biophys. 403 (2002) 1–15.
[17] K. Kawabata, T. Tanaka, S. Honjo, M. Kakumoto, A. Hara, H. Makita, N. Tatematsu, J. Ushida, H. Tsuda, H. Mori, Chemopreventive effect of dietary flavonoid morin on chemically induced rat tongue carcinogenesis, Int. J. Cancer 83 (1999) 381–386.
[18] S. Kilani-Jaziri, V. Frachet, W. Bhouri, K. Ghedira, L. Chekir-Ghedira, X. Ronot, Flavones inhibit the proliferation of human tumor cancer cell lines by inducing apoptosis, Drug Chem. Toxicol. 35 (2012) 1–10.
[19] G. Krauss, Biochemistry of Signal Transduction and Regulation, John Wiley & Sons, 2006.
[20] P.D. Kwong, M.L. Doyle, D.J. Casper, C. Cicala, S.A. Leavitt, S. Majeed, T.D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P.W.I.H. Parren, J. Robinson, D. Van Ryk, L. Wang, D.R. Burton, E. Freire, R. Wyatt, J. Sodroski, W.A. Hendrickson, J. Arthos, HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites, Nature 420 (2002) 678–682.
[21] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd edition, Springer, New York, 2011.
[22] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, I.D. Kuntz, DOCK 6: combining techniques to model RNA-small molecule complexes, Rna 15 (2009) 1219–1230.
[23] C.C. Lin, F.A. Melo, R. Ghosh, K.M. Suen, L.J. Stagg, J. Kirkpatrick, S.T. Arold, Z. Ahmed, J.E. Ladbury, Inhibition of basal FGF receptor signaling by dimeric Grb2, Cell 149 (2012) 1514–1524.
[24] Q.hua Lu, C.dan Ba, D.ying Chen, Investigating noncovalent interactions of rutin - serum albumin by capillary electrophoresis - frontal analysis, J. Pharm. Biomed. Anal. 47 (2008) 888–891.
[25] Y. Ma, A. Ge, W. Zhu, Y.-.N. Liu, N.-.F. Ji, W.-.J. Zha, J.-.X. Zhang, X.-.N. Zeng, M. Huang, Morin attenuates ovalbumin-induced airway inflammation by modulating oxidative stress-responsive MAPK signaling, Oxid. Med. Cell. Longev. (2016).
[26] K. Machida, B.J. Mayer, The SH2 domain: versatile signaling module and pharmaceutical target, Biochim. Biophys. Acta - Proteins Proteomics 1747 (2005) 1–25.
[27] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins †, J. Phys. Chem. B 102 (1998) 3586–3616.
[28] S. Maignan, J. Guilloteau, N. Fromage, B. Arnoux, J. Becquart, A. Ducruix, Crystal structure of the mammalian Grb2 adaptor, Science 268 (1995) 291–293 (80-.).
[29] M. Mohammadi, I. Dikic, A. Sorokin, W.H. Burgess, M. Jaye, J. Schlessinger, Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction, Mol. Cell. Biol. 16 (1996) 977–989.
[30] P. Morlacchi, F.M. Robertson, J. Klostergaard, J.S. McMurray, Targeting SH2 domains in breast cancer, Future Med. Chem. (2014).
[31] A. Papadopoulou, R.J. Green, R.A. Frazier, Interaction of flavonoids with bovine serum albumin: a fluorescence quenching study, J. Agric. Food Chem. 53 (2005) 158–163.
[32] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera - a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[33] V.V. Plotnikov, J.M. Brandts, L.N. Lin, J.F. Brandts, A new ultrasensitive scanning calorimeter, Anal. Biochem. 250 (1997) 237–244.
[34] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. Van Der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854.
[35] D. Raffa, B. Maggio, M.V. Raimondi, F. Plescia, G. Daidone, Recent discoveries of anticancer flavonoids, Eur. J. Med. Chem., Curr. Adv. Cancer Res. 142 (2017) 213–228.
[36] J. Rahuel, B. Gay, D. Erdmann, A. Strauss, C. García-Echeverría, P. Furet, G. Caravatti, H. Fretz, J. Schoepfer, M.G. Grütter, Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode, Nat. Struct. Biol. 3 (1996) 586.
[37] P.D. Ross, S. Subramanian, Thermodynamics of protein association reactions: forces contributing to stability, Biochemistry 20 (1981) 3096–3102.
[38] K. Sanches, I.P. Caruso, F.C.L. Almeida, F.A. Melo, The dynamics of free and phosphopeptide-bound Grb2-SH2 reveals two dynamically independent subdomains and an encounter complex with fuzzy interactions, Sci. Rep. 10 (2020) 13040.
[39] K. Sanches, Í.P. Caruso, F.C.L. Almeida, F.A. Melo, NMR assignment of free 1H, 15N and 13C-Grb2-SH2 domain, Biomol. NMR Assign (2019).
[40] K. Sanches, R.V.R. Dias, P.H. da Silva, M.A. Fossey, Í.P. Caruso, F.P. de Souza, L.C. de Oliveira, F.A. de Melo, Grb2 dimer interacts with Coumarin through SH2 domains: a combined experimental and molecular modeling study, Heliyon 5 (2019).
[41] R.J. Santen, R.X. Song, R. McPherson, R. Kumar, L. Adam, M.-.H. Jeng, W. Yue, The role of mitogen-activated protein (MAP) kinase in breast cancer, J. Steroid Biochem. Mol. Biol. 80 (2002) 239–256.
[42] W. Tamiru, E. Engidawork, K. Asres, Evaluation of the effects of 80% methanolic leaf extract of Caylusea abyssinica (fresen.) fisch. & Mey. on glucose handling in normal, glucose loaded and diabetic rodents, BMC Complement. Altern. Med. 12 (2012) 151.
[43] Z. Timsah, J. Berrout, M. Suraokar, C. Behrens, J. Song, J.J. Lee, C. Ivan, M. Gagea, M. Shires, X. Hu, C. Vallien, C.V. Kingsley, I. Wistuba, J.E. Ladbury, Expression pattern of FGFR2, Grb2 and Plcγ1 acts as a novel prognostic marker of recurrence recurrence-free survival in lung adenocarcinoma, Am. J. Cancer Res. 5 (2015) 3135–3148.

[44] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Eng. Des. Sel. 8 (1995) 127–134.
[45] W.H.J. Ward, G.A. Holdgate, 7 Isothermal titration calorimetry in drug discovery, Prog. Med. Chem. 38 (2001) 309–376.
[46] J. Yan, A.D. Kline, H. Mo, M.J. Shapiro, E.R. Zartler, The effect of relaxation on the epitope mapping by saturation transfer difference NMR, J. Magn. Reson. 163 (2003) 270–276.
[47] V. Zoete, M.A. Cuendet, A. Grosdidier, O. Michielin, SwissParam: a fast force field generation tool for small organic molecules, J. Comput. Chem. 32 (2011) 2359–2368.

Food Research International 158 (2022) 111515

Contents lists available at ScienceDirect

Food Research International

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

Aqueous mixtures of cornstarch and Pluronic® F127 studied by experimental and computational techniques

Monize A.B. Baldassa [a], Raphael V.R. Dias [a], Leandro C. Oliveira [a], Eloi Feitosa [b,*]

[a] Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University "Júlio de Mesquita Filho" (UNESP), São José do Rio Preto, SP 15054-000, Brazil
[b] Academic Department of Chemistry and Biology, Federal Technological University of Paraná (UTFPR), Curitiba 81280-340, Brazil

ARTICLE INFO

Keywords:
Cornstarch
Pluronic F127
Starch solubility
Inclusion complex

ABSTRACT

The possibility of interaction between cornstarch (CS) and amphiphilic molecules, such as the micelle-forming triblock copolymer Pluronic® F127 (F127), also known by Poloxamer 407, indicates that CS-F127 aqueous mixtures can regulate either the starch solubility or the copolymer micellization. Herein experimental and computational techniques were used to investigate CS-F127 aqueous mixtures aiming to highlight the role of these compounds on the molecular complexation. Dynamic light scattering results show that CS in water is highly polydisperse, while the F127 concentration and temperature influence the micellization process and the interaction with CS. Circular dichroism data of CS supernatants indicate the existence of small helical-like granules ($D_h \approx 800$ nm) in the CS-F127 mixed aqueous solutions at 25 °C. UV–Vis spectrophotometry shows a small absorption band around 267 or 275 nm characteristic of micelles, granules, or molecular complexes, while FTIR and X-ray diffractometry indicate negligible structural changes. Lugol iodine tests at 25 °C show that both the precipitate and supernatant in the mixtures undergo some structural changes also indicating molecular complexation. Molecular dynamic simulations show the formation of stabilized inclusion complexes (V-amylose), where the propylene oxide segment of the copolymer inside the amylose helix and the ethylene oxide branches facing the aqueous media. These results together reveal weak CS-F127 interactions, evidencing a small solubility of CS both in the absence and presence of F127 as a solubilizing agent. Furthermore, moderate CS amounts do not change the F127 micelle structure.

1. Introduction

Starch in general, and cornstarch (CS) in particular, is a natural polysaccharide composed mainly of amylose and amylopectin widely used as a staple food (Whistler et al., 1984). Studies of the properties of starch have promoted a variety of industrial applications (Janssen and Moscicki, 2009; Tabasum et al., 2019]. In the food industry, CS is widely used as gelling, thickener, or emulsifier (Lu et al., 2012; Cui et al., 2018). It has also been used in the manufacture of biodegradable films in the material industry (Pavoni et al., 2019; Herniou-Julien et al., 2019), as a binder and pore-forming element in the ceramic industry (Živcová-Vlčková et al., 2012), as adhesive in the cardboard industry (Chen et al., 2020), as an excipient or disintegrant in the pharmaceutical industry (Sadeghi et al., 2019), among others.

CS is considered insoluble in water at room temperature since most (if not all) of the material precipitates. Overall, there is no consensus on the actual solubility of CS in water at room temperature (Chen et al., 2019; He et al., 2014; Obadi and Xu, 2020; Ratnayake and Jackson, 2006). It is well known, however, that the increase in temperature favors the CS solubility in water. Mishra and Rai (2006) demonstrated that native starch is poorly solubilized in water at 20 °C, while at 70 °C it is solubilized at different extent depending on the origin (CS is 3.9% soluble at this temperature). A remarkable characteristic of starch is the gelatinization temperature (T_g), above it the granules become more flexible due to hydration of their inner structure (Fanta and Christianson, 1996). As a result, the precipitate solubilizes and the solution turns gel (Bertolini, 2010). For native CS, $T_g = 78.25$ °C (Mishra and Rai, 2006).

The understanding of the starch solubilization is relevant for scientific and industrial applications. In order to properly change the

* Corresponding author at: Universidade Tecnológica Federal do Paraná (UTFPR), Departamento Acadêmico de Química e Biologia, Rua: Dep. Heitor de Alencar Furtado, 5000, Bl. C, 81280-340 Curitiba, PR, Brazil.
E-mail address: eloifeitosa@utfpr.edu.br (E. Feitosa).

https://doi.org/10.1016/j.foodres.2022.111515
Received 2 March 2022; Received in revised form 31 May 2022; Accepted 13 June 2022
Available online 22 June 2022
0963-9969/© 2022 Elsevier Ltd. All rights reserved.

solubility or T_g, starch has been modified by the addition of chemical groups (Jane, 1992; Sweedman et al., 2013). However, complete starch solubility is attained only by heating, followed by gelatinization as an irreversible process. Below T_g, the starch solubilization may also find industrial application (Khan, 2019), for example, as drug encapsulation (Nallasamy et al., 2020). The addition of amphiphilic molecules, such as Pluronic® F127 (F127), to a starch solution, can be used to favor solubilization at temperatures below T_g through hydrophobic effects, hydrogen bonds, or van der Waals forces (Putseys et al., 2010). The enhancement of the CS solubility at room temperature may be an advance for the production of food, pharmaceutical, and chemical products.

Starch is a good receptor for small molecules as the amylose helical structure displays both the hydrophilic and hydrophobic domains, respectively outside and inside the helix, thus enabling the adsorption of hydrophilic and hydrophobic molecules via electrostatic or hydrophobic interactions (Petkova-Olsson et al., 2017; Qi and Tester, 2019). The hydrophobic cavity of the amylose helix contains C—H bonds between the planes of the rings (Jonhed and Järnström, 2006) and displays adjustable diameter to the size of the complexing agent (Biliaderis and Galloway, 1989). Furthermore, amylose inclusion complexes with lipids may also assume the V-amylose helical conformation (Carlson et al., 1979; Putseys et al., 2010; López et al., 2012; Schahl et al., 2020).

Pluronic® F127 (F127) is a thermoresponsive, linear, and flexible molecule consisting of a 65-unit central hydrophobic propylene oxide (PO) bound chemically to two 100-unit hydrophilic terminal segments of ethylene oxide (EO) (Barba et al., 2009) widely used in scientific research and industrial applications (Akash and Rehman, 2015). The knowledge of physico-chemical properties of F127 (or another copolymer) in solution is required for practical applications (Batrakova et al., 2003), for example, to improve the administration of poorly water-soluble drugs (Basak and Bandyopadhyay, 2013; Salama, 2021), or as an artificial skin for treating burns (Schmolka, 1972; Nalbandian et al., 1987). The FDA (Food and Drug Administration) agency considers F127 as an inactive ingredient in inhalation, solution, and syrup (Rowe et al., 2009). The low toxicity also allows F127 to be used in clinical applications (Giuliano et al., 2018). A typical F127 micelle diameter is about 20 nm (Alexandridis and Hatton, 1995). The critical micellar concentration (CMC) and the aggregation number of F127 micelles in water are dependent on temperature (Attwood et al., 1985; Bouchemal et al., 2009). At 25 °C, the CMC is approximately 0.7 wt% (Alexandridis et al., 1994), but it decreases with temperature (Bouchemal et al., 2009), while the aggregation number and size increase (Attwood et al., 1985; Wanka, et al., 1990; Linse and Malmsten, 1992; Bohorquez et al., 1999). F127 also presents a critical micelle temperature (CMT) which decreases with increasing the copolymer concentration (Alexandridis et al., 1994). Changes in F127 micelle structures (e.g., from spherical to cylindrical) can be obtained by increasing temperature or Pluronic concentration, or by adding a co-solute molecule or co-solvent (Chaibundit et al., 2010; Kim et al., 2013; Jalaal et al., 2017).

Computer simulations, such as molecular dymamic (MD) techniques, are powerful tools to model and elucidate structural organization of molecules in a solvent under the action of different interaction forces (Moraitakis et al., 2003). MD was used in this investigation to provide the helical conformation of amylose (Cheng et al., 2018; 2019), the main molecular component of CS, as well as its interaction with F127. The helical structure of amylose has also been demonstrated experimentally (Hsein-Chih and Sarko (1978; Gessler et al., 1999). Based on surface tension and quartz crystal microbalance results, Petkova-Olsson et al. (2017) reported the formation of inclusion complexes between modified potato starch and F127 at 30 °C (above CMT), but not at 20 °C (below CMT). Our MD simulation results presented here reveal the formation of inclusion complexes both below (25 °C) and above (35 °C), possibly because of differing allowed helical structures displayed by amylose (Hsein-Chih and Sarko, 1978). F127 also interacts with natural or synthetic lipids (Feitosa and Winnik, 2010; Torcello-Gómez et al., 2013; Wood et al., 2016) as well as with small molecules (Salay et al., 2018) to form molecular complexes suitable for drug delivery.

As an attempt to detect and control the CS solubility in water, it was used here F127 as a complexing agent. CS alone or mixed with F127 in aqueous solution precipitate, as well as selected solid precipitate samples, were investigated by experimental techniques, including dynamic light scattering (DLS), Ultraviolet–visible (UV–vis) spectrophotometry, X-ray diffractometry, circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy, lugol iodine tests, and also by MD simulations of the amylose component of CS in water in the absence and presence of F127. It was expected that in presence of F127, CS would raise the solubility and strengthen the potential for drug encapsulation or other industrial applications.

2. Materials and methods

2.1. Materials

The cornstarch (CS) used here was commercial (Maizena®, Unilever - Brazil), lot JC300 and 63B, purchased in an ordinary supermarket. Pluronic® F127 (molecular mass of about 12,600 Da) was purchased from Sigma-Aldrich (Saint Louis, USA) and used as received. Potassium bromide (KBr) supplied by Shimadzu Corporation (Kyoto, Japan) was of analytical grade. Lugol iodine was obtained by mixing 5% of I_2 (Conelgo Produtos Químicos Ltda., São Paulo, Brazil) and 10% of KI (Merck, Darmstadt, Germany) with 85% water. Purified water of Milli-Q quality (resistivity 18.2 MΩ·cm at 25 °C) was used in sample preparation. As the molecular mass of CS used is unknown, its concentration is expressed here as a percentage by mass wt%.

2.2. Sample preparation

All samples were prepared at room temperature (around 25 °C). Volumes of 15 mL of up to 8.0 wt% F127 aqueous samples were prepared by mixing the appropriate amount of this compound with water in a glass vial to attain the desired concentration, followed by a vigorous magnetic stirring to attain a homogeneous solution within around 1 h. CS (sample S1) and CS:F127 (sample V1) aqueous solutions (15 mL) at the desired concentrations of the compounds were prepared by adding, respectively water and F127 solution to an appropriate amount of CS previously weighed, followed by a magnetic stirring. Then, samples S1 and V1 were left standing during four days for precipitation. Then, 10 mL of sample S1 and V1 supernatant was taken (sample S2 and V2) and after six further days of standing, 5 mL of samples S2 and V2 supernatants were taken (sample S3 and V3) for analyses.

The precipitate from sample V1 was kept during 4 days in an oven at 30 °C for drying (solid sample V4). The aqueous samples were investigated by DLS, UV–Vis and CD spectrophotometry, while the solid and precipitate samples were investigated by FTIR and X-ray diffraction techniques.

It is worth noticing that during the storage time, structural degradation of the starch may occur due to microbial activities. For application purposes this degradation must be minimized by adding antimicrobial agents to the CS-containing samples (Muszanska et al. 2011). Such degradation effects were not regarded in the present study because of the very low CS composition in the supernatant samples, as well as the use of purified water as a solvent. Further, F127 may function as antibacterial and anti-adhesive (Valen and Bruzell 2018), thus enabling the minimization of the CS degradation process.

2.3. DLS measurements

The scattering equipment used (Brookhaven Instruments Corporation, Holtsville, USA) contain a BI 200SM goniometer, a digital autocorrelator, and a 35 mW He-Ne laser operating at the wavelength of 632.8 nm. Approximately 3 mL of aqueous sample in a cylindrical glass

cuvette was used for the DLS measurements. The measurements were made at the scattering angle of 90° and at temperatures of 15, 25, and 35 °C. The sample temperature was controlled by a circulating bath (Poly Science Temperature Control, model 91A0A11B). The data analysis was performed using the program OriginPro® supplied by the manufacturer.

2.4. UV–Vis spectrophotometry measurements

For the absorbance measurements, square quartz cuvettes of 1 cm optical length were used. A spectrophotometer, model Cary Bio 300 UV–Vis (Varian, Mulgrave, Australia), was used for the measurements in the wavelength range of 200–800 nm with water in the reference cuvette. In the data analysis, the absorbance of water was discounted from the sample signal.

2.5. CD spectrometry measurements

Square quartz cuvettes of 0.5 cm optical length were used to obtain the CD spectra using a JASCO CD spectropolarimeter, model J-815 (Providence, RI, USA). The scans were performed at the speed of 50 nm·min^{-1} and the resulting spectra were the average of 10 accumulations subtracted the average of 10 accumulations of the blank signal.

2.6. FTIR spectrometry measurements

For the FTIR experiments, pellets with diameter of approximately 1.3 cm of the materials investigated (solid or precipitate materials) were prepared by adding a small amount of KBr to the samples (KBr pellet method). The following samples were prepared: solid CS, F127, and CS:F127 (10:10 mg), and the precipitates of CS:F127 (sample V4) at 0.2:6 wt% and 1.2:1.2 wt%. A 80 kN force was applied to the KBr-containing solid or precipitate samples during 3 min at room temperature to form the pellets (Fig. 1). The FTIR spectra were obtained at 25 °C in the wavenumber range of 400–4000 cm^{-1} using a Shimadzu IRTracer-100 FTIR spectrometer (Columbia, MD, USA). Each spectrum was calculated from 64 scans with a resolution of 4 cm^{-1}. A background signal was removed from the spectra of the samples in order to subtract the signal from the atmospheric CO_2.

2.7. X-ray diffractometry measurements

Amorphous glass plates containing approximately 200 mg of material were mounted for the X-ray diffractometry measurements. The diffractograms were recorded using an X-ray diffractometer (MiniFlex II, Rigaku), at diffraction angle range of 3–50 deg, CuKα radiation wavelength of 1.54 Å, and scan rate of 2 deg·min^{-1}. The device was operated at electrical voltage and current of 30 kV and 15 mA, respectively.

2.8. Lugol iodine reagent test

The lugol iodine tests were performed for the following samples: pure water, CS at 0.05% in water (sample S1), F127 at 1.2% in water, CS:F127 at 0.05:1.2% (sample V1, V2, and V3). Approximately 20 μL of lugol iodine was added to 5 mL of these samples and the vials were manually shaken and left standing for visual analysis.

2.9. Molecular dynamics simulation

The computational simulation methods used here were based on previous studies of similar carbohydrate systems, force fields, and solvent models (Cheng et al., 2018; 2019). The initial structure of the amylose was obtained through the GLYCAM server (Kirschner et al., 2008) by setting it up in a helix form (V-type). The F127 molecular structure was built using the antechamber graphical tool (Wang et al., 2006), and its charge was assigned by selecting the AM1 with bond charge correction (AM1-BCC) model within the antechamber program in AmberTools18 (Case et al., 2018; Lee et al., 2018). The xLeap program in the Amber18 package was used to prepare the system. The force fields GLYCAM_06h (Lee et al., 2018) and the generalized AMBER (GAFF, version 1.5) (Wang et al., 2004) were used for the amylose and F127, respectively. Two different aqueous systems were built: one containing only amylose and the other containing both amylose and F127.

The MD simulations were carried out using the Amber18 package (Case et al., 2018; Lee et al., 2018) and analyzed through AmberTools. In all stages it was used the TIP3P water model and a time step of 2 fs. The SHAKE algorithm (Ryckaert et al., 1977) was applied to constrain all bond lengths and the box was built with a truncated octahedral box around the molecules (Case et al., 2018; Lee et al., 2018) with periodic boundary conditions (PBC). Two minimization steps were performed: the first one employing only the solvent molecules in the calculations, while in the second step the whole system was used. Both steps were performed using 50,000 steepest-descent and 5,000 conjugate gradient steps. Then, the system was coupled to a Berendsen barostat (Berendsen et al., 1984) and to a Langevin thermostat (Loncharich et al., 1992), respectively for pressure and temperature control. The systems were simulated at 298 and 308 K (~25 and 35 °C) and at 1 atm. As the results obtained were similar, only those at 298 K are presented here. The heating steps were performed during 1 ns run at constant volume, in the NVT (canonical) ensemble. The equilibration stage ended with a dynamic in the NPT (isothermal-isobaric) ensemble using isotropic pressure scaling and a relaxation time of 2 ps. The MD simulations were performed during 100 ns, and the trajectory position and energy were recorded every 100 ps.

3. Results and discussion

The semi-crystalline granules of cornstarch (CS) are insoluble in water and precipitate at temperatures below its gelatinization temperature (T_g) (Oates, 1997; Liu et al., 2010). Above T_g, the granules are disrupted and the molecules are dispersed in water as a gel. Below T_g, only a small fraction of CS is dissolved in water as a colloidal suspension, while most granules precipitate (Silva et al., 2017). Physical and

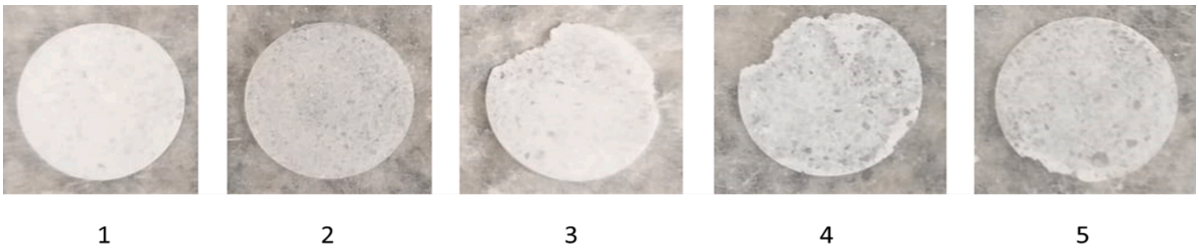


Fig. 1. Pellets for FTIR measurements formed with samples of solid CS (1) and F127 (2), V4 sample of CS:F127 0.2:6% (3), CS:F127 1.2:1.2% (4), and solid CS:F127 10:10 mg (5).

chemical protocols have been used to quantify and control the starch solubility, given by the ratio of the soluble mass to initial mass of starch (Silva et al., 2017). These authors reported a solubility of 0.1% for the CS in water. In order to verify the solubility of CS in water and the CS-F127 interactions below T_g, aqueous solutions, precipitates, and solid samples of the CS and F127 individual or mixed compounds were investigated by a number of experimental techniques and molecular dynamics (MD) simulations. The later showed inclusion complex formation of V-amylose type between the CS component amylose with F127, similar to the reported for modified potato starch and F127 (Petkova-Olsson et al., 2017).

3.1. Dynamic light scattering (DLS) results

Fig. 2a shows the D_h mean values of the structures present in the aqueous solutions of neat F127 at 6.0 wt%, supernatant samples S2 and S3 of neat CS at 0.2 wt%, and supernatants V2 and V3 of mixed CS:F127 at 0.2:6.0 wt%, at 15, 25, and 35 °C. Fig. 2b shows similar diagram, but for F127 and CS at 1.2 wt%.

According to Fig. 2a, at 15 °C (i.e., below CMT) 6.0 wt% of neat F127 assemble as pre-micelles with $D_h \approx 4.0$ nm, while at 25 and 35 °C (i.e., around and above CMT), they assemble as micelles with $D_h \approx 35$ and 17 nm, respectively. Samples S2 and S3 of neat CS at 1.2 wt% display granules with D_h in the range of 500–700 nm at the temperatures investigated.

At 15 °C, the samples V2 and V3 of CS:F127 at 0.2:6.0 wt% are dominated by granules of $D_h \approx 630$ nm and $D_h \approx 735$ nm, respectively, with a small population of pre-micelles with $D_h \approx 4.0$ nm. At 25 °C, surprisingly the mixed samples V2 and V3 are dominated by micelle-like structures with $D_h \approx 15$ and 40 nm. These results may indicate that CS granules are complexed with F127 monomers at 15 °C, while small micelle-like complexes with some compounds of CS are also formed at 25 and 35 °C.

Fig. 2b shows that at 1.2 wt%, neat F127 assembles as pre-micelles at 15 and 25 °C and as micelles with $D_h \approx 20$ nm at 35 °C. The assembly of F127 in water, as monomer, pre-micelle, or micelle, might influence the strength and type of the interaction with CS. For the samples S2 and S3 of neat CS at 1.2%, $D_h \approx 600$ nm regardless of the temperature. At 15 and 25 °C, the CS:F127 solutions at 1.2:1.2 wt% (samples V2 and V3) contain mainly granules with D_h in the range of 760–860 nm. At 35 °C, micelle-like structures with $D_h \approx 20$ nm also co-exist with the CS-F127 complexes. Such data also suggest that F127 monomers might be interacting with the granule without changing considerably the granule size, while the excess F127 molecules assemble as micelle structures.

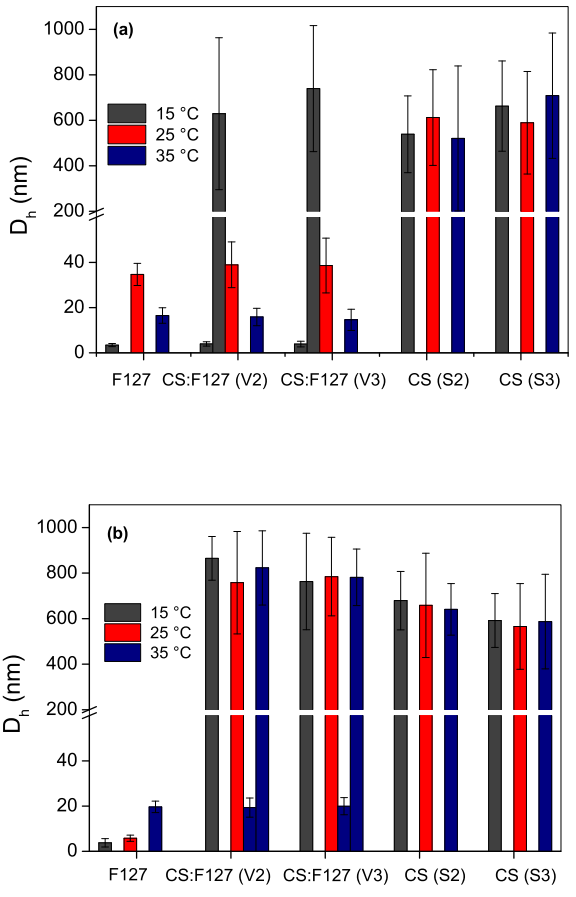


Fig. 2. Bar diagrams for hydrodynamic diameter (D_h) values of the assembly structures in the aqueous samples at 15, 25 and 35 °C: (a) F127 at 6.0 wt%, samples V2 and V3 of CS:F127 at 0.2:6.0 wt%, samples S2 and S3 of CS at 0.2 wt %, and (b) F127 at 1.2 wt%, samples V2 and V3 of CS:F127 at 1.2:1.2 wt%, and samples S2 and S3 of CS at 1.2 wt%.

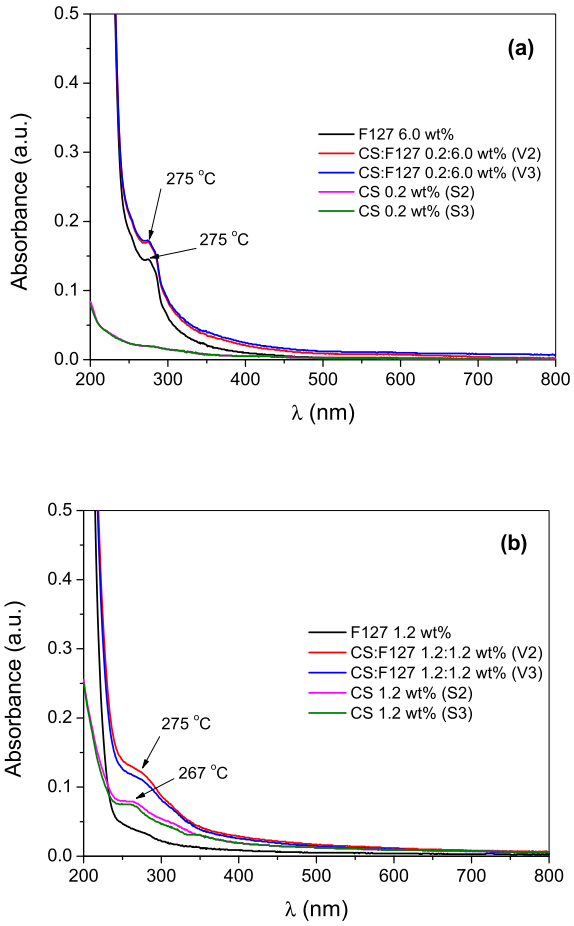


Fig. 3. UV–Vis absorbance spectra of the following aqueous samples at 25 °C: (a) F127 at 6.0 wt%, samples S2 and S3 of CS at 0.2 wt%, and samples V2 and V3 of CS:F127 at 0.2:6.0 wt%, and (b) F127 at 1.2 wt%, samples S2 and S3 of CS at 1.2 wt%, and samples V2 and V3 of CS:F127 at 1.2:1.2 wt%. Arrows point the band maxima.

3.2. UV–Vis spectrophotometry results

Fig. 3a and 3b show UV–Vis absorbance spectra of CS at 0.2 and 1.2 wt% (samples S2 and S3), F127 at 1.2 and 6.0 wt%, and mixtures of these compounds in water (samples V2 and V3). Accordingly, for all samples the absorbance is very low in the visible region and decreases smoothly to zero, indicating low scattering.

In the UV region, the 1.2 wt% CS and 6.0 wt% F127 aqueous solutions display absorbance bands around 267 and 275 nm, respectively; at the lowest concentrations of these compounds (0.2 and 1.2 wt%), these bands are not discernible. Salaheldin (2018) also reported similar absorbance spectrum for 1 wt% CS prepared in hot water (>50 °C). Samples V2 and V3 of CS:F127 mixtures at 0.2:6.0 and 1.2:1.2 wt% display the CS band around 275 nm.

Owing to the little effect of the CS-F127 mixture on the absorbance of the individual compounds, these data together suggest physical interactions (e.g., steric and/or hydrophobic interactions) between these compounds. Furthermore, the very similar results for samples S2 and S3 and for V2 and V3 indicate that these samples are quite analogous, that is, samples S2 and V2 present no precipitate.

3.3. CD spectroscopy results

The CD spectroscopy was used to detect helical structures in the CS-containing aqueous samples. Fig. 4 shows the CD spectra of neat CS (samples S2 and S3) and F127 at 1.2 wt%, and of CS:F127 mixture at 1.2:1.2 wt% (samples V2 and V3). As expected, the CD spectrum of 1.2 wt% F127 solution is nearly zero with no optical spectrum variation, indicating micellar or pre-micellar assembly. The CD spectra for the samples S2 and S3 of CS at 1.2 wt%, on the other hand, display a small negative ellipticity of approximately −0.5 mdeg within the λ range of 235–210 nm, indicating the presence of small helical granules of amylose. The CD spectra for the samples V2 and V3 of CS:F127 at 1.2:1.2 wt% mixture also display a negative band of approximately −0.7 mdeg within the λ range of 240–210 nm. Such a spectral deviation resembles that for 2-hexanone complexed with amylose hydroxypropilated (Wulff and Kubik, 1992), indicating formation of CS-F127 inclusion complexes (Putseys et al., 2010; Obiro et al., 2012; Petkova-Olsson et al., 2017).

3.4. FTIR spectrometry results

Fig. 5 shows the FTIR spectra in the wavenumber range of 1800–400 cm^{-1} for solid CS, F127, and CS:F127 (10:10 mg) mixture, and for the precipitates of CS:F127 at 0.2:6.0 and 1.2:1.2 wt% aqueous mixtures (see the sample images in Fig. 1). In order to ascertain the existence of molecular complexes in the supernatant of the CS-F127 aqueous samples, it was compared some peaks characteristic of solid CS or F127 with those of the precipitates.

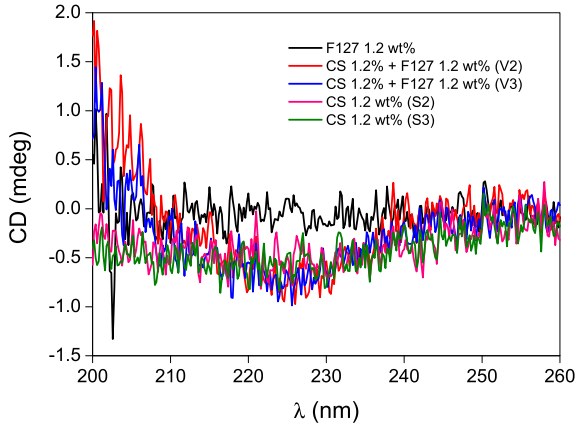


Fig. 4. CD spectra of neat F127 at 1.2 wt%, samples S2 and S3 of CS at 1.2 wt%, and mixed samples V2 and V3 of CS:F127 at 1.2:1.2 wt% aqueous solutions at 25 °C. Purified water was used as reference.

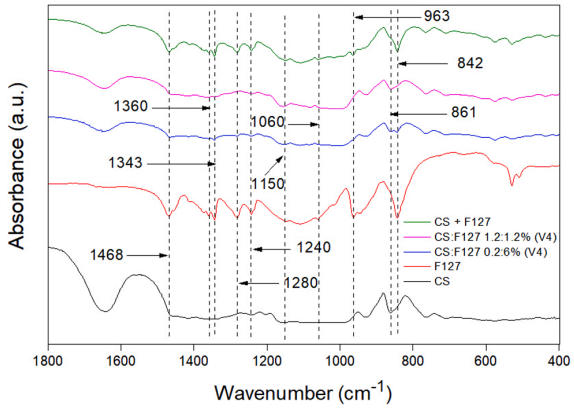


Fig. 5. FTIR spectra of the following samples at 25 °C: solid CS, F127, and CS:F127 (10:10 mg), and precipitates (sample V4) of CS:F127 at 0.2:6.0 and 1.2:1.2 wt% aqueous mixtures.

The peak at 1468 cm^{-1} is due to the bending vibration of the CH_2 group of F127, while the peaks at 1343 and 1281 cm^{-1} are assigned to the angular deformation of CH_2 from F127 (Su et al., 2002). Accordingly, the peaks are more intense for the solid F127 than for the CS:F127 solid mixtures, which, in turn, is more intense than the spectra for the precipitates. Since there is no significant shift of these peaks, the change in the peak intensity might be related to the amount of F127 molecules in the aggregates. It may also indicate that the F127 molecules are in the supernatant as free monomers, micelles, or CS-F127 complexes. Furthermore, the peak at 861 cm^{-1} is characteristic of the C–O–C stretching in the CS structure (Abdullah et al., 2018; Kibar and Us, 2014). A similar but weaker peak is seen in the spectra of samples V4 of CS:F127 at 0.2:6.0 and 1.2:1.2 wt%, probably as a result of the sample dilution. In the solid CS:F127 (10:10 mg) sample, this peak is overlapped with that at 842 cm^{-1} characteristic of F127.

Other peaks including those at 1343 and 1360 cm^{-1} assigned to the wagging vibrations of the CH_2 groups of F127 (Branca et al., 2018), appear less intense in the solid CS:F127 mixture, but they are not discernible in the precipitate (V4) samples. Similar behavior was also observed for the peaks at 1240 and 1280 cm^{-1} for twisting vibrations of the CH_2 groups of F127, at 1150, 1060 cm^{-1} for C—O—C stretching vibrations, and at 962 and 946 cm^{-1} assigned to CH_2 rocking bands (Branca et al., 2018).

These FTIR results ensure the presence of F127 and CS both in the precipitate as well as in the supernatant samples. However, they do not allow one to precise whether the Pluronic molecules in solution are free or associated as micelles or complexed with granules.

3.5. X-Ray diffractometry (XRD) results

The X-ray diffractograms for the solid CS, F127, and CS:F127 (10:10 mg) mixtures, and for the precipitates (sample V4) of the CS:F127 at 0.2:6.0 and 1.2:1.2 wt% aqueous samples are shown in Fig. 6. Cheetham and Tao (1998) demonstrated that the X-ray diffractogram profile of a solid native CS sample depends on the amylose content. Comparison of the diffractogram profile of our solid commercial CS (Fig. 6) with those reported by Cheetham and Tao, it can be deduced that our CS sample contains about 28% of amylose, indicating that our CS is of crystalline A-type. Some differences in the profiles of these diffractograms might be due to the commercial vs native origin of the CS samples. Similarly, the

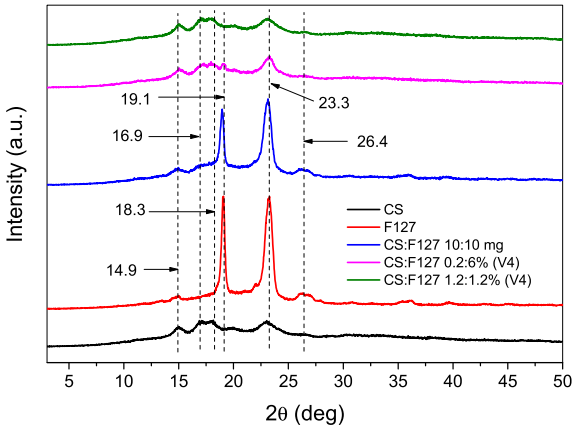


Fig. 6. X-Ray diffractograms for the following samples at 25 °C: solid CS, F127, and precipitates (sample V4) of CS:F127 (0.2:6 and 1.2:1.2 wt%) aqueous solutions.

characteristic peaks for solid F127 resemble those reported in the literature (Parmar et al., 2011; Gülsün et al., 2012).

The diffractogram for the CS:F127 (10:10 mg) mixture seems to be an overlap of the diffractograms for neat CS and F127 with the characteristic bands being less intense because of the lower amounts of the compounds in the mixture. The diffractogram profile for the precipitate (sample V4) of the CS:F127 (1.2:1.2 wt%) aqueous solution is mostly like the diffractogram for the solid CS, while the diffractogram profile for the precipitate (sample V4) of the CS:F127 (0.2:6.0 wt%) aqueous solution displays bands characteristic of both CS and F127. These findings indicate that the structure of the CS-F127 complex formed may depend on the relative amount of the compounds.

The band at 19.1 deg characteristic for F127 is present in the diffractogram for the precipitate sample V4 of CS:F127 (0.2:6.0 wt%), indicating that a fraction of F127 precipitates most probably complexed with CS granules. Such band is not distinguished for the precipitate of CS:F127 (1.2:1.2 wt%) possibly because of the smaller amount of CS in this sample. The characteristic peaks for CS display negligible shifts in the diffractograms for the CS:F127 mixtures. The increase in the intensity of the band at 23.3 deg may indicate CS-F127 complexation as F127 also presents a band near 23.3 deg. However, without knowing the precise concentrations of the components in the samples, such a statement cannot be confirmed here.

Thus, similar to the FTIR results, and despite the evidence of the CS-F127 interaction and complexation, it is not possible yet to identify whether monomeric or micellar F127 is found inside or outside of the granules. Our computer simulation (Section 3.7) points to the F127 monomer interaction with the amylose component of CS.

3.6. Lugol iodine tests

Fig. 7 shows images of selected samples after addition of 20 μL lugol iodine (LI) into 5 mL aqueous solutions of CS, F127, and CS:F127 mixtures, labelled L1-L6, as indicated in the figure text. Accordingly, the reference LI solution displays a characteristic dark reddish color. The bluish color of the LI-containing starch samples indicates the interaction of the triiodide ions with the amylose helices with the formation of the amylose-iodide complex (Pfannemüller and Ziegast, 1982).

A yellowish solution (sample L1) was obtained by dilution of LI in water. In sample L2 (LI in sample S1 of CS at 0.05 wt%) the precipitate looks bluish while the supernatant is clear indicating the presence of granules in the precipitate. The clear supernatant might be due to the small amount of solubilized CS. Sample L3 (LI in neat 1.2 wt% F127 solution) displays an intense yellow color, which might be due to the presence of hydroxyl groups in F127 that interact with water molecules, thus reducing the interaction of LI with water.

Samples L4, L5, and L6 contain LI, respectively in samples V1, V2, and V3 of CS:F127 (0.05:1.2 wt%) solution. The precipitate of sample L4 initially displayed a light red–purple color, suggesting a small interaction of iodide with the amylopectin chains, according to Shen et al. (2013) studies on similar systems. The L4, L5, and L6 samples remained colorless, probably because the iodine was bound to the CS-F127 complexes rather than directly to the amylopectine structure. Alternatively, the F127 monomers might bind physically to the iodine sites inside the granules, as suggested by the computer simulation results presented next.

3.7. Molecular dynamics (MD) results

A computational approach was employed with the purpose of evaluating the amylose representative structures in the absence and presence of F127. For the simulations, the molecular structures of amylose and F127 were shortened proportionally from their actual size in order to work consistently with the computational resources available. The amylose-F127 complex was built by fitting the hydrophobic block of the copolymer into the carbohydrate spiral (V-amylose) and with the hydrophilic blocks facing the solvent.

Fig. 8a and 8b show frames of the amylose molecular structure in water, respectively in the absence and presence of F127, for different simulation periods up to 100 ns. Accordingly, the temporal evolution reveals that the F127-free amylose helical structure is unfolded (Fig. 8a), while in presence of F127, the amylose inclusion complex with F127 in the V-amylose structure (Fig. 8b), is preserved in the interaction region and in its surrounding (region highlighted with dashed lines). Far from the copolymer, the granule behaves as an unfolding helix characteristic of amylose in the absence of F127. Fig. 8c was obtained from the MD trajectories, where the calculations considered the initial frame as the reference structure. As the simulations were performed during a reasonable time, these results are comparable to each other (based on the molecular size of the system), thus the deviation (RMSD) values are higher for the system containing only amylose than for the F127-containing system. Therefore, these computational analyses support the hypothesis raised after the experiments that F127 induces and maintains the V-amylose conformation.

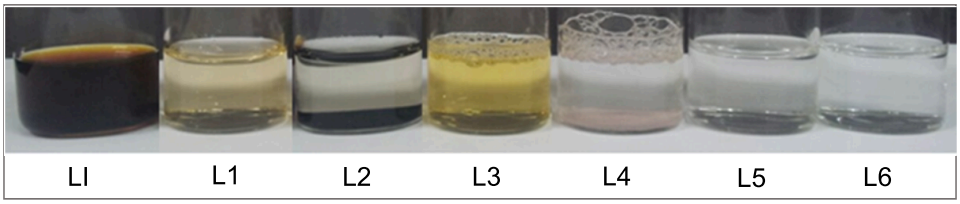


Fig. 7. Pictures taken at 25 °C of lugol iodine (LI) solution, and of LI diluted in water (L1), in CS (0.05 wt%, S1) (L2), and in aqueous solutions of F127 at 1.2% in the absence (L3) and in presence of CS at 0.05%, samples V1 (L4), V2 (L5), and V3 (L6).

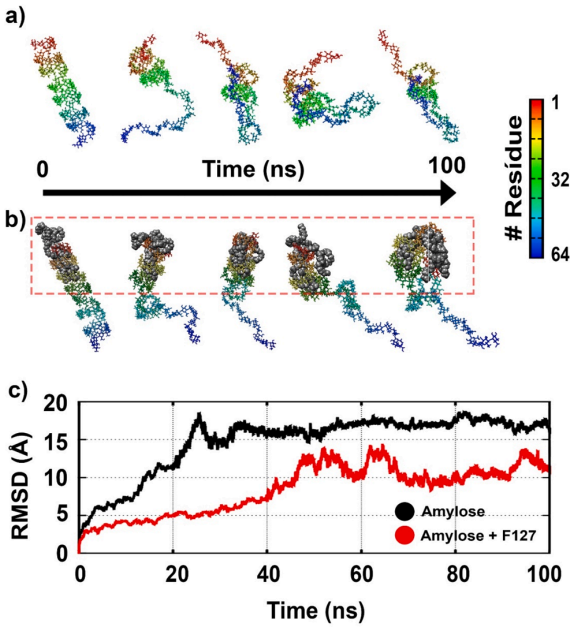


Fig. 8. Simulation evaluation of amylose in water at 298 K (~25 °C) in the absence and presence of F127. The temporal evolution of the amylose (a) and amylose + F127 (b) based on the simulations were represented by the licorice drawing method for carbohydrates and VDW for the copolymer at equally spaced times along the simulation. The color bar was built up based on the residue number along the amylose, going from red (residue 1) to blue (residue 64). The variation related to the initial structure and the dynamical behavior was evaluated by using the RMSD (c). RMSD values calculated using the CPPTRAJ tool available in the Amber18 package (Case et al., 2018; Lee et al., 2018). The images were constructed using the VMD software (Humphrey et al., 1996). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions

By using experimental and computational techniques it was demonstrated that cornstarch (CS) displays low solubility in water and complexed with Pluronic® F127 (F127) below the CS gelatinization temperature (T_g). These results together demonstrated that in aqueous mixtures of these compounds, CS-F127 complexes are present both in the supernatant and precipitate. The helical structure of CS granules with a characteristic size of about 800 nm was certified by CD spectra and DLS. MD simulations established that amylose inclusion complexes of the V-amylose type are formed with the PO hydrophobic central part of the copolymer inside and the hydrophilic EO segments outside the amylose helix. Furthermore, excess F127 in solution may be either as free monomers, pre-micelles, micelles, or complexed with the amylopectin component of CS. CS-F127 complexes may be used as carriers of drugs or nutrients for pharmaceutical, nutritional, or aesthetic purposes. Hopefully, the present study may be further extended by using more advanced techniques in order to quantify and control the solubilization of starch in an aqueous solution below T_g aiming for specific applications.

Authors contributions

All the co-authors of our paper are researchers. L.C. Oliveira and me are professors, and M.A.B. Baldassa and R.V.R. Dias are young PhD researchres. Dr. Oliveira and Dias were involved with the computer simulation studies, while Dr. Baldassa and me tere involved with the experimental procedures. All of us were involved in both the discussion of the results and the writing of the manuscript. Responding to your request, we would say that the co-authors, besides me, were involved in the following: Supervision, Investigation, Methodology, Writing-original-draft, and Writing-review-editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The computational work was supported by São Paulo Research Foundation (FAPESP), grant 2016/08753–6. Computational evaluations were performed at the Center for Scientific Computing (NCC/Grid-UNESP) of the São Paulo State University (UNESP) and CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em São Paulo), project UNICAMP/FINEP-MCTII. LCO was supported by National Council for Scientific and Technological Development (CNPq), grant 442352/2014-0. EF thanks CNPq for PQ grant 309346/2019-3, and MABB thanks Coordination for the Improvement of Higher Education Personnel (CAPES) and FAPESP (grant 2016/08753–6) for scholarship.

References

Abdullah, A. H. D., Chalimah, S., Primadona, I., & Hanantyo, M. H. G. (2018). Physical and chemical properties of corn, cassava, and potato starchs. *IOP Conference Series: Earth and Environmental Science, 160*, Article 012003. https://doi.org/10.1088/1755-1315/160/1/012003

Akash, M. S. H., & Rehman, K. (2015). Recent progress in biomedical applications of Pluronic (PF127): Pharmaceutical perspectives. *Journal of Controlled Release, 209*, 120–138. https://doi.org/10.1016/j.jconrel.2015.04.032

Alexandridis, P., Holzwarth, J. F., & Hatton, T. A. (1994). Micellization of Poly(ethylene oxide)-Poly(propylene oxide)-Poly(ethylene oxide) Triblock Copolymers in Aqueous Solutions: Thermodynamics of Copolymer Association. *Macromolecules, 27*(9), 2414–2425. https://doi.org/10.1021/ma00087a009

Alexandridis, P., & Hatton, T. A. (1995). Poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide) block copolymer surfactants in aqueous solutions and at interfaces: Thermodynamics, structure, dynamics, and modeling. *Colloids and Surfaces A: Physicochemical and Engineering Aspects, 96*(1–2), 1–46. https://doi.org/10.1016/0927-7757(94)03028-x

Attwood, D., Collett, J., & Tait, C. (1985). The micellar properties of the poly (oxyethylene)-poly(oxypropylene) copolymer Pluronic F127 in water and electrolyte solution. *International Journal of Pharmaceutics, 26*(1–2), 25–33. https://doi.org/10.1016/0378-5173(85)90197-8

Barba, A. A., d'Amore, M., Grassi, M., Chirico, S., Lamberti, G., & Titomanlio, G. (2009). Investigation of Pluronic© F127-Water solutions phase transitions by DSC and dielectric spectroscopy. *Journal of Applied Polymer Science, 114*(2), 688–695. https://doi.org/10.1002/app.30586

Basak, R., & Bandyopadhyay, R. (2013). Encapsulation of Hydrophobic Drugs in Pluronic F127 Micelles: Effects of Drug Hydrophobicity, Solution Temperature, and pH. *Langmuir, 29*(13), 4350–4356. https://doi.org/10.1021/la304836e

Batrakova, E. V., Li, S., Alakhov, V. Y., Miller, D. W., & Kabanov, A. V. (2003). Optimal Structure Requirements for Pluronic Block Copolymers in Modifying P-glycoprotein Drug Efflux Transporter Activity in Bovine Brain Microvessel Endothelial Cells. *Journal of Pharmacology and Experimental Therapeutics, 304*(2), 845–854. https://doi.org/10.1124/jpet.102.043307

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics, 81*(8), 3684–3690. https://doi.org/10.1063/1.448118

Bertolini, A. C. (2010). *Starch: Characterization, properties, and applications*. Boca Raton: CRC Press.

Biliaderis, C. G., & Galloway, G. (1989). Crystallization behavior of amylose-v complexes: Structure-property relationships. *Carbohydrate Research, 189*, 31–48. https://doi.org/10.1016/0008-6215(89)84084-4

Bohorquez, M., Koch, C., Trygstad, T., & Pandit, N. (1999). A Study of the Temperature-Dependent Micellization of Pluronic F127. *Journal of Colloid and Interface Science, 216*(1), 34–40. https://doi.org/10.1006/jcis.1999.6273

Branca, C., Khouzami, K., Wanderlingh, U., & D'Angelo, G. (2018). Effect of intercalated chitosan/clay nanostructures on concentrated pluronic F127 solution: A FTIR-ATR, DSC and rheological study. *Journal of Colloid and Interface Science, 517*, 221–229. https://doi.org/10.1016/j.jcis.2018.02.004

Bouchemal, K., Agnely, F., Koffi, A., & Ponchel, G. (2009). A concise analysis of the effect of temperature and propanediol-1, 2 on Pluronic F127 micellization using isothermal titration microcalorimetry. *Journal of Colloid and Interface Science, 338*(1), 169–176. https://doi.org/10.1016/j.jcis.2009.05.075

Carlson, T.-L.-G., Larsson, K., Dinh-Nguyen, N., & Krog, N. (1979). A study of the amylose-monoglyceride complex by Raman spectroscopy. *Starch, 31*(7), 222–224. https://doi.org/10.1002/star.19790310703

Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E. I. I. I., Cruzeiro, V. W. D., et al. (2018). *AMBER 2018*. San Francisco: University of California.

Chaibundit, C., Ricardo, N. M. P. S., Ricardo, N. M. P. S., Muryn, C. A., Madec, M.-B., Yeates, S. G., & Booth, C. (2010). Effect of ethanol on the gelation of aqueous solutions of Pluronic F127. *Journal of Colloid and Interface Science, 351*(1), 190–196. https://doi.org/10.1016/j.jcis.2010.07.02

Cheetham, N. W., & Tao, L. (1998). Variation in crystalline type with amylose content in maize starch granules: An X-ray powder diffraction study. *Carbohydrate Polymers, 36*(4), 277–284. https://doi.org/10.1016/s0144-8617(98)00007-1

Chen, H., Chen, L., Dang, X., Shan, Z., Dai, R., & Wang, Y. (2019). The structure and properties of granular cold-water-soluble starch by a NaOH/urea aqueous solution. *International Journal of Biological Macromolecules, 141*, 732–737. https://doi.org/10.1016/j.ijbiomac.2019.08.226

Chen, L., Xiong, Z., Xiong, H., & Din, Z. (2020). Investigating the Structure and Self-assembly Behavior of Starch-g-VAc in Starch-based Adhesive by Combining NMR Analysis and Multi-scale Simulation. *Carbohydrate Polymers, 246*, Article 116655. https://doi.org/10.1016/j.carbpol.2020.116655

Cheng, L., Feng, T., Zhang, B., Zhu, X., Hamaker, B., Zhang, H., & Campanella, O. (2018). A molecular dynamics simulation study on the conformational stability of amylose-linoleic acid complex in water. *Carbohydrate Polymers, 196*, 56–65. https://doi.org/10.1016/j.carbpol.2018.04.102

Cheng, L., Zhu, X., Hamaker, B. R., Zhang, H., & Campanella, O. H. (2019). Complexation process of amylose under different concentrations of linoleic acid using molecular dynamics simulation. *Carbohydrate Polymers, 216*, 157–166. https://doi.org/10.1016/j.carbpol.2019.04.013

Cui, S., Li, M., Zhang, S., Liu, J., Sun, Q., & Xiong, L. (2018). Physicochemical properties of maize and sweet potato starches in the presence of cellulose nanocrystals. *Food Hydrocolloids, 77*, 220–227. https://doi.org/10.1016/j.foodhyd.2017.09.037

Fanta, G. F., & Christianson, D. D. (1996). Starch-hydrocolloid composites prepared by steam jet cooking. *Food Hydrocolloids, 10*(2), 173–178. https://doi.org/10.1016/s0268-005x(96)80032-4

Feitosa, E., & Winnik, F. M. (2010). Interaction between Pluronic F127 and dioctadecyldimethylammonium bromide (DODAB) vesicles studied by differential scanning calorimetry. *Langmuir, 26*(23), 17852–17857. https://doi.org/10.1021/la102603a

Gessler, K., Uson, I., Takaha, T., Krauss, N., Smith, S. M., Okada, S., Sheldrick, G. M., & Saenger, W. (1999). V-Amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaose). *Proceedings of the National Academy of Sciences, 96*(8), 4246–4251. https://doi.org/10.1073/pnas.96.8.4246

Giuliano, E., Paolino, D., Fresta, M., & Cosco, D. (2018). Mucosal applications of poloxamer 407-based hydrogels: An overview. *Pharmaceutics, 10*(3), 159. https://doi.org/10.3390/pharmaceutics10030159

Gülsün, T., Budak, Ç., Vural, İ., Sahin, S., & Öner, L. (2012). Preparation and characterization of nimesulide containing nanocrystal formulations. *Pharmaceutical Development and Technology, 18*(3), 653–659. https://doi.org/10.3109/10837450.2012.663390

He, S., Qin, Y., Walid, E., Li, L., Cui, J., & Ma, Y. (2014). Effect of ball-milling on the physicochemical properties of maize starch. *Biotechnology Reports, 3*, 54–59. https://doi.org/10.1016/j.btre.2014.06.004

Herniou-Julien, C., Mendieta, J. R., & Gutiérrez, T. J. (2019). Characterization of biodegradable/non-compostable films made from cellulose acetate/corn starch blends processed under reactive extrusion conditions. *Food Hydrocolloids, 89*, 67–79. https://doi.org/10.1016/j.foodhyd.2018.10.024

Hsein-Chih, H. W., & Sarko, A. (1978). The double-helical molecular structure of crystalline b-amylose. *Carbohydrate Research, 61*(1), 7–25. https://doi.org/10.1016/s0008-6215(00)84463-8

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics, 14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Jalaal, M., Cottrell, G., Balmforth, N., & Stoeber, B. (2017). On the rheology of Pluronic F127 aqueous solutions. *Journal of Rheology, 61*(1), 139–146. https://doi.org/10.1122/1.4971992

Jane, J.-L. (1992). Preparation and food applications of physically modified starches. *Trends in Food Science & Technology, 3*, 145–148. https://doi.org/10.1016/0924-2244(92)90169-w

Janssen, L. P. B. M., & Moscicki, L. (2009). *Thermoplastic Starch: A Green Material for Various Industries*. Weinheim: Wiley-VCH.

Jonhed, A., & Järnström, L. (2006). The Interaction Between Surfactants and 2-Hydroxy-3-(N, N-dimethyl-N-dodecylammonium)-propyloxy Starches. *Starch/Stärke, 58*(11), 561–571. https://doi.org/10.1002/star.200500481

Kibar, E. A. A., & Us, F. (2014). Evaluation of structural properties of cellulose ether-corn starch based biodegradable films. *International Journal of Polymeric Materials and Polymeric Biomaterials, 63*(7), 342–351. https://doi.org/10.1080/00914037.2013.845190

Khan, Z. (2019). Encapsulation of silver nanoparticles into the helix of water soluble starch and their sensing properties. International Journal of Biological Macromolecules. *International Journal of Biological Macromolecules, 136*, 165–176. https://doi.org/10.1016/j.ijbiomac.2019.05.012

Kim, T.-H., Han, Y.-S., Jang, J.-D., & Seong, B.-S. (2013). SANS study on self-assembled structures of Pluronic F127 triblock copolymer induced by additives and temperature. *Journal of Applied Crystallography, 47*(1), 53–59. https://doi.org/10.1107/s1600576713030094

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. Journal of Computational Chemistry, 29*(4), 622–655. https://doi.org/10.1002/jcc.20820

Lee, T.-S., Cerutti, D. S., Mermelstein, D., Lin, C., LeGrand, S., Giese, T. J., Roitberg, A., Case, D. A., Walker, R. C., & York, D. M. (2018). GPU-accelerated molecular dynamics and free energy methods in Amber18: Performance enhancements and new features. *Journal of Chemical Information and Modeling, 58*(10), 2043–2050. https://doi.org/10.1021/acs.jcim.8b00462

Linse, P., & Malmsten, M. (1992). Temperature-dependent micellization in aqueous block copolymer solutions. *Macromolecules, 25*(20), 5434–5439. https://doi.org/10.1021/ma00046a048

Liu, P., Yu, L., Wang, X., Li, D., Chen, L., & Li, X. (2010). Glass transition temperature of starches with different amylose/amylopectin ratios. *Journal of Cereal Science, 51*(3), 388–391. https://doi.org/10.1016/j.jcs.2010.02.007

Loncharich, R. J., Brooks, B. R., & Pastor, R. W. (1992). Langevin dynamics of peptides: The frictional dependence of isomerization rates ofN-acetylalanyl-N'-methylamide. *Biopolymers, 32*(5), 523–535. https://doi.org/10.1002/bip.360320508

López, C. A., de Vries, A. H., & Marrink, S. J. (2012). Amylose folding under the influence of lipids. *Carbohydrate Research, 364*, 1–7. https://doi.org/10.1016/j.carres.2012.10.007

Lu, J., Luo, Z., & Xiao, Z. (2012). Effect of lysine and glycine on pasting and rheological properties of maize starch. *Food Research International, 49*(1), 612–617. https://doi.org/10.1016/j.foodres.2012.06.038

Mishra, S., & Rai, T. (2006). Morphology and functional properties of corn, potato and tapioca starches. *Food Hydrocolloids, 20*(5), 557–566. https://doi.org/10.1016/j.foodhyd.2005.01.001

Moraitakis, G., Purkiss, A. G., & Goodfellow, J. M. (2003). Simulated dynamics and biological macromolecules. *Reports on Progress in Physics, 66*(3), 383–406. https://doi.org/10.1088/0034-4885/66/3/203

Muszanska, A. K., Busscher, H. J., Herrmann, A., van der Mei, H. C., & Norde, W. (2011). Pluronic–lysozyme conjugates as anti-adhesive and antibacterial bifunctional polymers for surface coating. *Biomaterials, 32*(26), 6333–6341. https://doi.org/10.1016/j.biomaterials.2011.05.016

Nalbandian, R. M., Henry, R. L., Balko, K. W., Adams, D. V., & Neuman, N. R. (1987). Pluronic F-127 gel preparation as an artificial skin in the treatment of third-degree burns in pigs. *Journal of Biomedical Materials Research, 21*(9), 1135–1148. https://doi.org/10.1002/jbm.820210907

Nallasamy, P., Ramalingam, T., Thajuddin, N., Shanmuganathan, R., Pugazhendhi, A., & Natarajan, S. (2020). Polyherbal drug loaded starch nanoparticles as promising drug delivery system: Antimicrobial, antibiofilm and neuroprotective studies. *Process Biochemistry, 92*, 355–364. https://doi.org/10.1016/j.procbio.2020.01.026

Oates, C. G. (1997). Towards an understanding of starch granule structure and hydrolysis. *Trends in Food Science & Technology, 8*(11), 375–382. https://doi.org/10.1016/s0924-2244(97)01090-x

Obadi, M., & Xu, B. (2020). Review on the physicochemical properties, modifications, and applications of starches and its common modified forms used in noodle products. *Food Hydrocolloids, 112*, Article 106286. https://doi.org/10.1016/j.foodhyd.2020.106286

Obiro, W. C., Ray, S. S., & Emmambux, M. N. (2012). V-amylose Structural Characteristics, Methods of Preparation, Significance, and Potential Applications. *Food Reviews International, 28*(4), 412–438. https://doi.org/10.1080/87559129.2012.660718

Parmar, B., Mandal, S., Petkar, K. C., Patel, L. D., & Sawant, K. K. (2011). Valsartan loaded solid lipid nanoparticles: Development, characterization, and In vitro and ex vivo evaluation. *International Journal of Pharmaceutical Sciences and Nanotechnology, 4*(3), 1483–1490. https://doi.org/10.37285/ijpsn.2011.4.3.7

Pavoni, J. M. F., Luchese, C. L., & Tessaro, I. C. (2019). Impact of acid type for chitosan dissolution on the characteristics and biodegradability of cornstarch/chitosan based films. *International Journal of Biological Macromolecules, 138*, 693–703. https://doi.org/10.1016/j.ijbiomac.2019.07.089

Petkova-Olsson, Y., Altun, S., Ullsten, H., & Järnström, L. (2017). Temperature effect on the complex formation between Pluronic F127 and starch. *Carbohydrate Polymers, 166*, 264–270. https://doi.org/10.1016/j.carbpol.2017.02.012

Pfannemüller, B., & Ziegast, G. (1982). Characterization of structural changes of amylose-iodine solutions by circular dichroism. *International Journal of Biological Macromolecules, 4*(1), 9–17. https://doi.org/10.1016/0141-8130(82)90004-6

Putseys, J. A., Lamberts, L., & Delcour, J. A. (2010). Amylose-inclusion complexes: Formation, identity and physico-chemical properties. *Journal of Cereal Science, 51*(3), 238–247. https://doi.org/10.1016/j.jcs.2010.01.011

Qi, X., & Tester, R. F. (2019). Starch granules as active guest molecules or microorganism delivery systems. *Food Chemistry, 271*, 182–186. https://doi.org/10.1016/j.foodchem.2018.07.177

Ratnayake, W. S., & Jackson, D. S. (2006). Gelatinization and Solubility of Corn Starch during Heating in Excess Water: New Insights. *Journal of Agricultural and Food Chemistry, 54*(10), 3712–3716. https://doi.org/10.1021/jf0529114

Rowe, R. C., Sheskey, P. J., & Quinn, M. E. (2009). *Handbook of Pharmaceutical Excipients* (6th ed.). London: Pharmaceutical Press.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics, 23*(3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5

Sadeghi, M., Hemmati, S., Salehi, R., Solhi, M., Ghorbani, M., & Hamishehkar, H. (2019). Leucine-grafted starch as a new superdisintegrant for the formulation of

domperidone tablets. *Journal of Drug Delivery Science and Technology, 50*, 136–144. https://doi.org/10.1016/j.jddst.2019.01.021

Salaheldin, H. I. (2018). Corrigendum: Optimizing the synthesis conditions of silver nanoparticles using corn starch and their catalytic reduction of 4-nitrophenol. *Advances in Natural Sciences: Nanoscience and Nanotechnology, 9*(3), Article 039501. https://doi.org/10.1088/2043-6254/aacf11

Salama, A. H. (2021). Pluronic F127 and its applications. *Pharmacologyonline, 2*, 1393–1403.

Salay, L. C., Prazeres, E. A., Huachaca, N. S. M., Lemos, M., Piccoli, J. P., Sanches, P. R. S., Cilli, E. M., Santos, R. S., & Feitosa, E. (2018). Molecular interactions between Pluronic F127 and the peptide tritrpticin in aqueous solution. *Colloid and Polymer Science, 296*, 809–817. https://doi.org/10.1007/s00396-018-4304-0

Schahl, A., Réat, V., & Jolibois, F. (2020). Structures and NMR spectra of short amylose-lipid complexes. Insight using molecular dynamics and DFT quantum chemical calculations. *Carbohydrate Polymers*. , Article 115846. https://doi.org/10.1016/j.carbpol.2020.115846

Schmolka, I. R. (1972). Artificial skin I. Preparation and properties of pluronic F-127 gels for treatment of burns. *Journal of Biomedical Materials Research, 6*(6), 571–582. https://doi.org/10.1002/jbm.820060609

Shen, X., Bertoft, E., Zhang, G., & Hamaker, B. R. (2013). Iodine binding to explore the conformational state of internal chains of amylopectin. *Carbohydrate Polymers, 98*(1), 778–783. https://doi.org/10.1016/j.carbpol.2013.06.050

Silva, A. C., Silva, E. M. S., Peres, A. E. C., & Sousa, D. N. (2017). Temperature influence in cornstarch gelatinization for froth flotation. *REM - International Engineering Journal, 70*(2), 231–235. https://doi.org/10.1590/0370-44672016700085

Su, Y. L., Wang, J., & Liu, H. Z. (2002). FTIR Spectroscopic Investigation of Effects of Temperature and Concentration on PEO−PPO−PEO Block Copolymer Properties in Aqueous Solutions. *Macromolecules, 35*(16), 6426–6431. https://doi.org/10.1021/ma0105284

Sweedman, M. C., Tizzotti, M. J., Schäfer, C., & Gilbert, R. G. (2013). Structure and physicochemical properties of octenyl succinic anhydride modified starches: A review. *Carbohydrate Polymers, 92*(1), 905–920. https://doi.org/10.1016/j.carbpol.2012.09.040

Tabasum, S., Younas, M., Zaeem, M. A., Majeed, I., Majeed, M., Noreen, A., Iqbalc, M. N., & Zia, K. M. (2019). A review on blending of corn starch with natural and synthetic polymers, and inorganic nanoparticles with mathematical modeling. *International Journal of Biological Macromolecules, 122*, 969–996. https://doi.org/10.1016/j.ijbiomac.2018.10.092

Torcello-Gómez, A., Maldonado-Valderrama, J., Jódar-Reyes, A. B., & Foster, T. J. (2013). Interactions between Pluronics (F127 and F68) and Bile Salts (NaTDC) in the Aqueous Phase and the Interface of Oil-in-Water Emulsions. *Langmuir, 29*(8), 2520–2529. https://doi.org/10.1021/la3044335

Valen, H., Bruzell, E., 2018. Antibacterial and anti-adhesive properties of Pluronic® F127 solutions in the absence or presence of photoactivation. Niom, April 27th 2018, in https://niom.no/antibacterial-and-anti-adhesive-properties-of-pluronic-f127-solutions-in-the-absence-or-presence-of-photoactivation/, access in May 24th, 2022.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry, 25*(9), 1157–1174. https://doi.org/10.1002/jcc.20035

Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling, 25*, 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005

Wanka, G., Hoffmann, H., & Ulbricht, W. (1990). The aggregation behavior of poly-(oxyethylene)-poly-(oxypropylene)-poly-(oxyethylene)-block-copolymers in aqueous solution. *Colloid & Polymer Science, 268*(2), 101–117. https://doi.org/10.1007/bf01513189

Whistler, R. L., BeMiller, J. N., & Paschall, E. F. (1984). *Starch: Chemistry and technology* (2nd ed.). Orlando: Academic Press.

Wood, I., Martini, M. F., Albano, J. M. R., Cuestas, M. L., Mathet, V. L., & Pickholz, M. (2016). Coarse grained study of pluronic F127: Comparison with shorter co-polymers in its interaction with lipid bilayers and self-aggregation in water. *Journal of Molecular Structure, 1109*, 106–113. https://doi.org/10.1016/j.molstruc.2015.12.073

Wulff, G., & Kubik, S. (1992). Circular dichroism and ultraviolet spectroscopy of complexes of amylose. *Carbohydrate Research, 237*, 1–10. https://doi.org/10.1016/S0008-6215(92)84228-K

Živcová-Vlčková, Z., Locs, J., Keuper, M., Sedlářová, I., & Chmelíčková, M. (2012). Microstructural comparison of porous oxide ceramics from the system Al2O3–ZrO2 prepared with starch as a pore-forming agent. *Journal of the European Ceramic Society, 32*(10), 2163–2172. https://doi.org/10.1016/j.jeurceramsoc.2012.02.005

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB Article

Csk αC Helix: A Computational Analysis of an Essential Region for Conformational Transitions

Published as part of The Journal of Physical Chemistry virtual special issue "Jose Onuchic Festschrift".

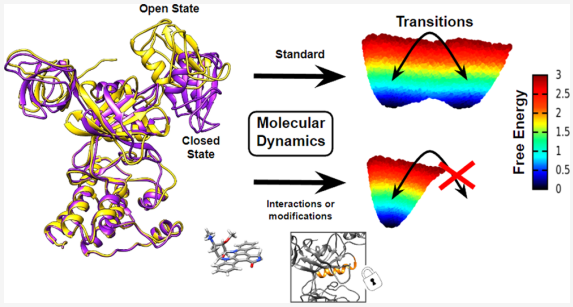
Raphael Vinicius Rodrigues Dias, Carolina Tatiani Alves Ferreira, Patricia Ann Jennings, Paul Charles Whitford, and Leandro Cristante de Oliveira*

Cite This: https://doi.org/10.1021/acs.jpcb.2c05408 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

Downloaded via UNIV ESTADUAL PAULISTA on December 13, 2022 at 17:09:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ABSTRACT: Conformational changes are an essential feature for the function of some dynamic proteins. Understanding the mechanism of such motions may allow us to identify important properties, which may be directly related to the regulatory function of a protein. Also, this knowledge may be employed for a rational design of drugs that can shift the balance between active and inactive conformations, as well as affect the kinetics of the activation process. Here, the conformational changes in carboxyl-terminal Src kinase, the major catalytic repressor to the Src family of kinases, was investigated, and it was proposed as a functionally related hypothesis. A Cα Structure-Based Model (Cα-SBM) was applied to provide a description of the overall conformational landscape and further analysis complemented by detailed molecular dynamics simulations. As a first approach to Cα-SBM simulations, reversible transitions between active (closed) and inactive (open) forms were modeled as fluctuations between these two energetic basins. It was found that, in addition to the interdomain Carboxyl-terminal SRC Kinase (Csk) correlated motions, a conformational change in the αC helix is required for a complete conformational transition. The result reveals this as an important region of transition control and domain coordination. Restrictions in the αC helix region of the Csk protein were performed, and the analyses showed a direct correlation with the global conformational changes, with this location being propitious for future studies of ligands. Also, the Src Homology 3 (SH3) and SH3 plus Src Homology 2 (SH2) domains were excluded for a direct comparison with experimental results previously published. Simulations where the SH3 was deleted presented a reduction of the transitions during the simulations, while the SH3−SH2 deletion vanishes the Csk transitions, corroborating the experimental results mentioned and linking the conformational changes with the catalytic functionality of Csk. The study was complemented by the introduction of a known kinase inhibitor close to the Csk αC helix region where its consequences for the kinetic behavior and domain displacement of Csk were verified through detailed molecular dynamics. The findings describe the mechanisms involving the Csk αC helix for the transitions and also support the dynamic correlation between SH3 and SH2 domains against the Csk lobes and how local energetic restrictions or interactions in the Csk αC helix can play an important role for long-range motions. The results also allow speculation if the Csk activity is restricted to one specific conformation or a consequence of a state transition, this point being a target for future studies. However, the αC helix is revealed as a potential region for rational drug design.



INTRODUCTION

Conformational changes are an important feature of some regulatory proteins and molecular motors, playing an essential role in living cells. The comprehension of the mechanisms involved in this process has allowed the understanding of several biological processes.[1−3] One interesting target for these studies is the Carboxyl-terminal SRC Kinase (Csk), a catalytic repressor responsible for phosphorylation, cell-growth control, and differentiation, being the main down regulator of Src Family Kinases (SFKs).[4−7] The proper functioning of Csk is crucial to cell health, as illustrated by the fact that the loss of function of Csk leads to constitutive activation of SFKs, accompanied by defects in embryonic development.[8] On the other hand, the gain of function of Csk can readily downregulate SFK-mediated cell

Received: July 30, 2022
Revised: November 22, 2022

© XXXX American Chemical Society

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

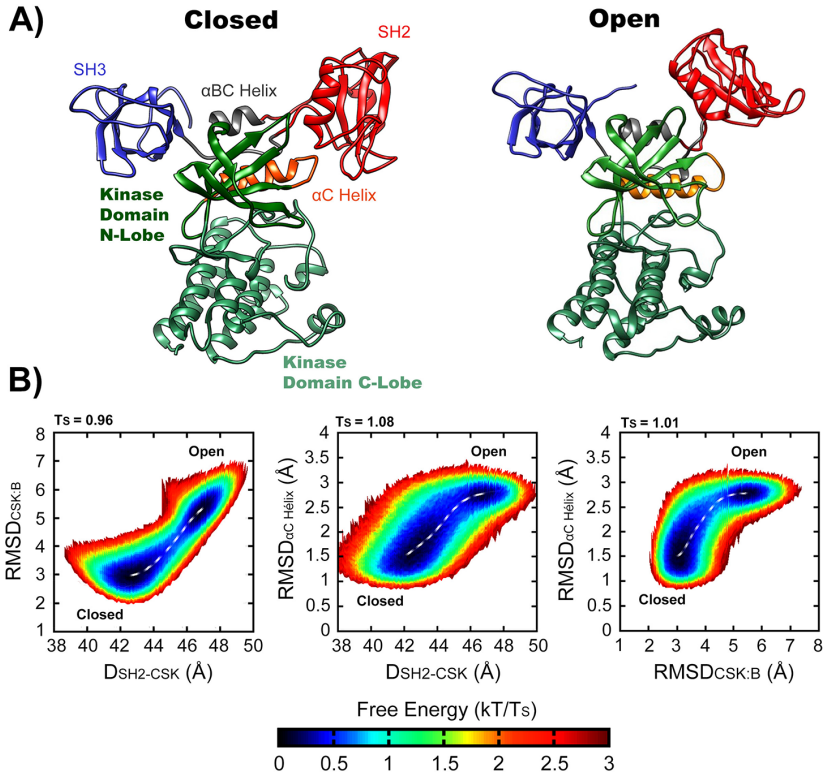


Figure 1. Csk conformational change analysis. (A) Csk cartoon representation of the closed (left) and open (right) conformations from the crystallographic structure (PDB ID: 1K9A, chains B and C, respectively), colored in order to highlight domains and specificity. SH3, SH3−SH2 linker, SH2, αC helix, small kinase lobe (N-Lobe), and large kinase lobe (C-Lobe) are in blue, gray, red, orange, green bean, and green forest, respectively. (B) Free energy surface of Csk calculated from Cα-SBM simulations using three different reaction coordinates: distance between the SH2 domain and C-terminal Lobe Kinase ($D_{SH2-CSK}$), Root Mean Square Deviation for the full structure ($RMSD_{CSK:B}$), and RMSD of the Csk αC helix ($RMSD_{\alpha CHelix}$), both RMSDs using the chain B (closed form) as reference. The free energy values follow the colorbar scale presented below from 0 (black) to 3 (dark red) normalized by the transition state temperature (T_S). The dashed line represents the lowest energy path between the two minimums.

signaling, which results in pathological consequences such as colon cancer.[9]

Csk is a 450-residue protein composed of four domains: Src Homology 2 (SH2), Src Homology 3 (SH3), and kinase small (Csk N-Lobe) and kinase large (Csk C-Lobe) lobes. The X-ray crystallographic structure (2.5 Å resolution) presents 6 chains where 2 of them (chains C and D) are highlighted for the more distant positioning of the SH2 domain, among other differences, assumed as inactive (open form), and 4 more compact ones (chains A, B, E, and F) are considered as active (closed form), identified in the Protein Data Bank by the code 1K9A.[10]

However, small-angle X-ray scattering analysis revealed a dynamic equilibrium between representative ensembles of conformations more extended and collapsed in solution[11] that can be shifted in the presence of ligands such as adenylyl imidodiphosphate (AMP-PNP), a nonhydrolyzable analogue of adenosine 5′-triphosphate (ATP) and adenosine 5′-diphosphate (ADP).[12]

All four domains have been shown to be required for the proper function of Csk,[10,13−16] though the atomistic understanding of the role of each domain on the global function is not fully understood. In many tyrosine protein kinases, the balance between active and inactive conformations is largely affected by the phosphorylation state of the C-terminal tyrosine tail,[17] presumably by directly stabilizing a particular relative positioning of the domains.

Figure 1A shows a cartoon illustration of the open and closed Csk forms based on the crystallographic structure, identifying domains and important local regions. When compared with other Src Family members, Csk lacks a C-terminal phosphorylation site and activation/inactivation is often induced by cofactor interactions that are distal to the active site and do not appear to induce large-scale domain rearrangements.[6,18,19] Accordingly, Csk must possess some form of internal communication channels between these distal sites and the ATP-binding site through modest shifts in either local configuration or dynamics.

Due to the biological importance of Csk, many studies have focused on exploring the interplay between the various domains of Csk and catalytic activity.[14,15,20−23] These studies have indicated important regions and residues that appear to be coupled to the functional properties of Csk. The importance of the noncatalytic domains (SH2 and SH3) is evidenced, for example, in domain deletion studies, which show the kinase activity reduced by 70% and 96% with the deletion of SH3 and SH2, respectively.[14,18,24] Still, the mutation of SER78, which is found in the SH2 domain, to Asp results in a reduction of the activity by more than 90%[14] under certain chemical scenarios.

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Some experiments using the deuterium incorporation tax to the wild-type and PHE183 mutants[15] reveal that αC helix and SH2-kinase linker are components that regulate the conformational change by the modulation between the SH2 domain and kinase domain. The same technique was applied to investigate the activation by CBP peptide, which increases the catalytic activity and suggests a mechanism of regulation.[15] NMR spectra obtained from oxidized and reduced forms of the SH2 domain showed that a particular disulfide bond can decrease the catalytic activity when formed, explained by the increasing strain around the binding site region and αC helix in SH2.[25] For all cases, normal-mode analysis showed how the flexibility of different regions is affected by mutations or CBP binding.[15] Despite this set of results showing the importance of communication between all domains present in Csk, a detailed description of these mechanisms can be exploited through computational evaluations, identifying the importance of specific regions for conformational changes in Csk and the development of new products, for example, the rational drug design for desirable features.

METHODS

Cα Structure-Based Model and Dual Basin Construction. Since the identification of essential elements for the mechanism's comprehension of conformational changes in Csk is desired, the Cα Structure-Based Model (Cα-SBM) is a suitable approach for the study. SBMs are versatile because they allow efficient conformational searching and as a consequence the employment of statistical mechanics for the analysis. They have been used for several purposes as the prospection of protein mechanisms,[26−28] conformational transitions,[1,29] or structural characterization when combined with experimental data.[11,12,30,31]

In this study, protein amino acids are represented by rigid spheres centered on the α carbon position.[32] The potential energy of the system is based on the geometrical characteristics of the native structure (Γ_0) compared to the current conformation (Γ) throughout the molecular dynamics simulation and given by

$$V(\Gamma, \Gamma_o) = \sum_{\text{bonds}} \varepsilon_r (r - r_o)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_o)^2 + \sum_{\text{backbone}} \varepsilon_\phi \left\{ [1 - \cos(\phi - \phi_o)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_o))] \right\} + \sum_{\text{contacts}} \varepsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{\text{non-contacts}} \varepsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \quad (1)$$

where r_{ij} is the distance between atom i and j and d_{ij} represents the distance between i and j in which the interaction energy is minimum. The instantaneous bond lengths, bond angles, and dihedral angles are given by r, θ, and ϕ, respectively, and r_0, θ_0, and ϕ_0 are the corresponding values in the native structure. The energetic prefactors used were: $\epsilon_r = 100\ \epsilon_C$, $\epsilon_\theta = 20\ \epsilon C$, and $\epsilon_\phi = \epsilon_{NC} = \epsilon_C$, where ϵ_C is equal to 1 unit (in reduced units) with contact interaction pairs defined by the Shadow Contact Map algorithm.[33] Structure-based models, in general, are widely employed for the study of protein mechanisms, and they are detailed in several references.[28,32,34]

The crystallographic configurations[10] and Small Angle X-ray Scattering analysis of Csk in solution[11,12] suggest the energy landscape of Csk does not have a single well, being composed of at least two basins. Since the inactive form (open) is highly degenerate and the main target is to reach the active form (closed) through conformational transitions, the use of the forms present in the crystal structure is acceptable. Thus, the evaluation of conformational changes involved the chains B and C, the closed and open forms, respectively, from the Csk crystal structure (PDB ID: 1K9A),[10] where missing residues were added using PROFIX.[35] However, individual dynamics of each chain potential [eq 1] are unable to access the other state; i.e., Cα-SBM molecular dynamics of chain B does not access the native structure of C and vice versa in any temperature (see Figure S3). Therefore, a hybrid potential (aka *Dual-Gō*) was constructed following the procedures employed by Whitford et al.,[1] combining the contacts of both conformations in a unique contact map. Exclusive contact pairs in closed form with distances 50% greater than in the open structure are selected and included in the open structure contact map, allowing by a fine-tuning the evaluation of conformational transitions between both forms. Consequently, the energy prefactors of the contacts (ϵ_C) added are multiplied by a constant ($2.1 \times \epsilon_C$), setting up the system for the conformational changes at an adequate range of temperatures between glass (T_g) and unfolding (T_u) (details in Figure S4). The energy function construction method assumes that these unique contacts selected from the closed form are responsible to drive the open conformation to the closed through the fine-tuning of the energetic prefactor. Nevertheless, it is important to mention that this approach requires conformations similar enough to have the essential contacts that drive one form to another present. Simulations where SH3 or SH2−SH3 domains were deleted employed this same adjusted potential, preserving only the applicable force field for the remaining residues.

Setup for SBM Simulations and Mimetic Construction of Inhibitors. The input files for SBM were generated by the SMOG web server[36] and modified when needed (see Cα Structure-Based Model and Dual Basin Construction). Molecular dynamics (MD) simulations were performed using GROMACS 4.5.7[37,38] where a temperature bath was coupled by Langevin Dynamics with a constant equal to 1 ps, as implemented elsewhere.[34] Simulations to evaluate conformational changes started with the protein in its open conformation in a total of 5×10^9 steps for each temperature evaluated and using a time step of 0.0005 time units. The configurations and energies were recorded every 5000 steps. Three different reaction coordinates were used to follow the conformational transitions between the two states: Distance between the SH2 and C-terminal Lobe Kinase domains ($D_{SH2-CSK}$), RMSD of the complete protein using the native structure in the closed form as a reference ($RMSD_{CSK:B}$), and RMSD of the αc helix located in the N-terminal Lobe Kinase ($RMSD_{\alpha C\ helix}$), also using as reference structure the closed form. The closed conformation (PBD ID: 1K9A:B) is given as the Csk active form;[10] thereby, it was chosen as a reference for comparative purposes (dihedral angles and RMSD values). The analyses were performed using a GROMACS provided tool, and the free energy profiles were calculated through the Weighted Histogram Analysis Method (WHAM).[39] The WHAM package was obtained from SMOG web server and for the calculations used 300 bins for each input

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

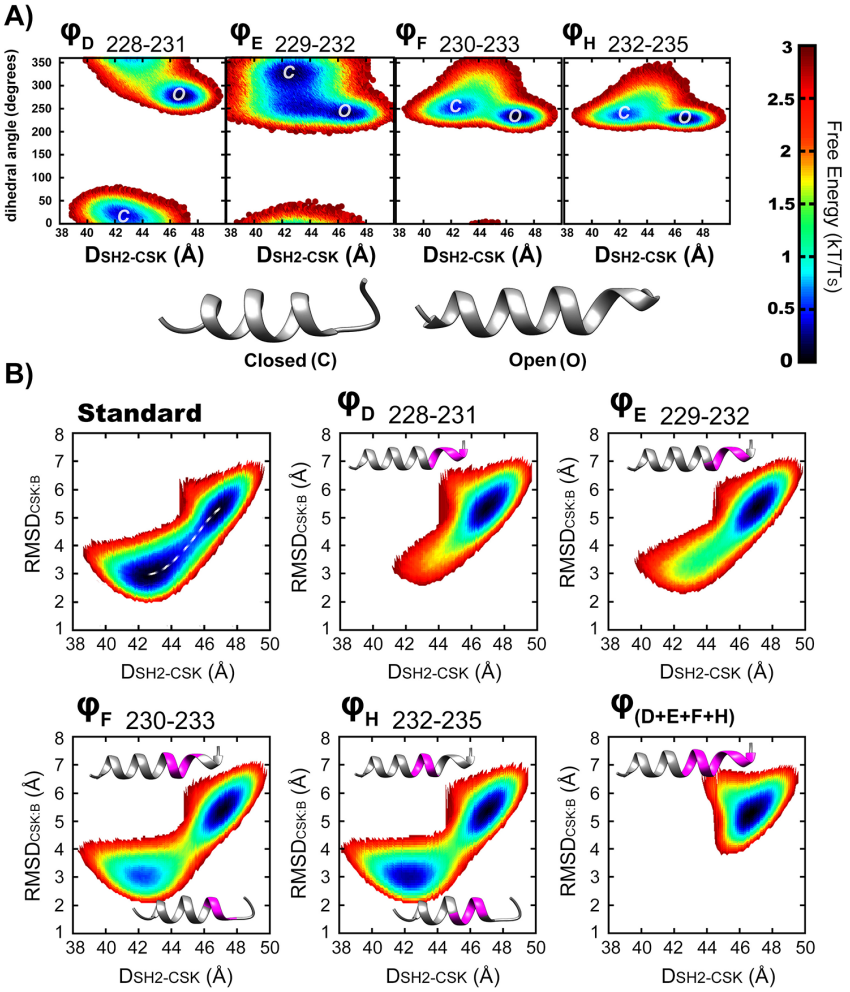


Figure 2. Effect of conformational changes under the viewpoint of Csk αC helix dihedral angles. (A) Free energy surfaces using the specifically indicated dihedral angles as reaction coordinates. (B) Influence of angular restrictions in located αC helix regions. The label "standard" is a free energy surface reference where dihedral angular restrictions were not imposed, followed by the specifically hardened angles. Disturbed dihedrals via an increase of pre-energy factors were highlighted in the cartoon representation in purple.

and a tolerance for the convergence of 1×10^{-3} in free energy variation.

The Csk−Staurosporine complex was constructed based on the crystallographic information available at PDB using the identification code 3D7T,[40] chain A (reinforced by PDB ID: 1QPJ, lymphocyte-specific protein tyrosine kinase (Lck) in complex with staurosporine), that lacks the SH2 and SH3 domains, through STAMP structural alignment[41] with the Csk structure (PDB ID: 1K9A). The staurosporine (STU) interaction effect is introduced implicitly using a harmonic potential $\left(\frac{1}{2} \times \epsilon_C \times (r - r_0)^2\right)$ where the residues i and j are determined by the distance from the residues to the inhibitor. If two residues are up to 6 Å apart from the same atom of the staurosporine, then they are considered interacting in the mimetic potential.

Detailed Molecular Dynamics Simulations. MD simulations using explicit water and ions were performed to evaluate the consistence of the mimetic construction of the inhibitor and complement the analysis with an inhibitor. These detailed MD simulations were performed with GROMACS 4.5.7 using CHARMM27 force field[42] with the standard setup for the protein residues and TIP3P water model. The ligand parametrization was calculated using the Web server Swiss-param from the Swiss Institute of Bioinformatics.[43] The MD minimization was done using 50000 steepest-descent steps and 5000 conjugate gradient steps, both without position restraints. The equilibration was performed in two steps with 100 ps, initially with and then followed without position restrictions for the protein and ligand atoms, when applicable. Initial atom velocities were based on the Maxwell−Boltzmann distribution and the MD simulations for 100 ns with an integration step of 2 fs at 300 K, salt concentration of 0.15 M for the net neutralization, and 1 atm. The trajectory positions and energies were recorded every 100 ps. The simulation used Parrinello−Rahman barostat,[44] Berendsen thermostat,[45] and LYNCS method to constrain all hydrogen bonds[46] and a simulation box built with the edges of the box with 10 Å from the

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

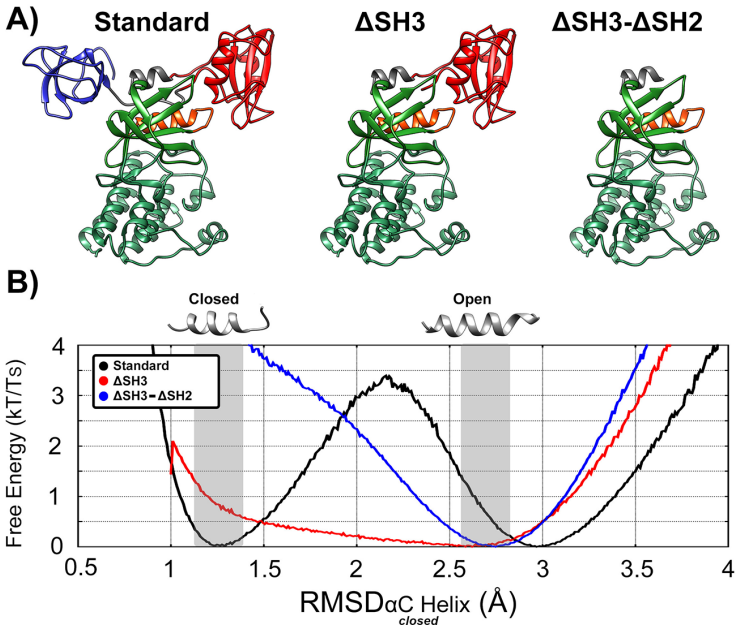


Figure 3. Effects of SH3 and SH3−SH2 domain deletions. (A) Cartoon representations of Csk tridimensional structure (left), Csk without SH3 domain (center), and Csk without SH3−SH2 domains (right), labeled by standard, ΔSH3, and ΔSH3−ΔSH2, respectively. (B) Free energy profiles based on the $RMSD_{\alpha C\ helix}$ (chain B as reference) reaction coordinate to standard, ΔSH3, and ΔSH3−ΔSH2.

protein surface and periodic boundary conditions. MD analyses were done using tools available in the GROMACS package.

RESULTS

Evaluation of Conformational Changes through Structure-Based Model Simulations. Previous studies have emphasized the importance of the Csk αC helix;[47−49] thus, initially, the Cα-SBM accuracy was evaluated to capture this feature. The free energy surfaces using Cα-SBM simulations with dual-conformational potential (hybrid potential) are presented in Figure 1B, where different reaction coordinates were employed, the distance between the SH2 domain and C-terminal Lobe Kinase ($D_{SH2-CSK}$), Csk RMSD ($RMSD_{CSK:B}$), and Csk αC helix RMSD ($RMSD_{\alpha C\ helix}$), where these last two parameters are related to the closed form (CSK:B). The analysis of these profiles reinforce how large movements, as described by the reaction coordinates $D_{SH2-CSK}$ and $RMSD_{CSK:B}$, are related to localized regions as the Csk αC helix. The free energy profiles under a unique coordinate reaction are also available in Figure S1 for viewing and quantification purposes.

Extensive searching of essential regions for functional transition control was performed for the simulations (data not shown). The analysis indicated the αC helix as the most promising region, where a detailed evaluation is presented in Figure 2A. The αC helix consists of 19 residues (from 225 to 243) that were divided into 16 dihedrals angles for calculation purposes. Each dihedral angle was used as a reaction coordinate to evaluate how specific residues are correlated with the $D_{SH2-CSK}$ values during the simulations. Since the relationship between the global coordinate reactions ($RMSD_{CSK:B}$ and $D_{SH2-CSK}$) and local values ($RMSD_{\alpha C\ helix}$ and dihedral angles) was verified, as presented in Figures 1B and 2A, this consideration as an indirect coordinate reaction is suitable. The dihedral angles were sequentially labeled ϕ_A, ϕ_B, ..., and ϕ_P, corresponding to angles formed by the residues 225−228, 226−229, ..., and 240−243. Among the 16 dihedrals analyzed (see Figure S2), four of them had a considerable barrier in the transition state on Csk conformational transitions, corresponding to the angles ϕ_D, ϕ_E, ϕ_F, and ϕ_H formed by the residues 228−231, 229−232, 230−233, and 232−235, respectively, detached in Figure 2A. The first two mentioned angles, ϕ_D and ϕ_E, have distinct values to the open and closed states properly characterizing these forms, while the two others, ϕ_F and ϕ_H, have basins with close values. Here, these close angular values can be understood as thermal fluctuations not effectively acting as a control region for conformational transitions. Thus, this small region involving ϕ_D and ϕ_E undergoes a high strain energy during the transitions and can work as a trigger for the shift between the open and closed states.

To evaluate the hypothesis of the Csk αC helix works as a region of control, an increase in the energetic prefactor of these dihedrals ($5 \times \epsilon\phi$) was made in order to stiffen the four dihedral torsions and probe their impact on the conformational changes. In Figure 2B, the free energy surfaces for the case using $RMSD_{\alpha C\ helix}$ versus $D_{SH2-CSK}$ are presented. Since the interference imposed in the dihedral angles affects its usage as a reaction coordinate for the analysis, a new one was employed. From the results, it is possible to note that the restriction imposed to the angles ϕ_E, ϕ_F, and ϕ_H solely is unfavorable for the conformational transitions by the increase of the transition state barrier. With the restriction imposed to ϕ_D exclusively, transitions become more limited, and when all of them (ϕ_D, ϕ_E, ϕ_F and ϕ_H) are restricted, they just do not occur. It is important to clarify that nonequal populations of the basin in these profiles are due to the protein starting to unfold with the increase of the temperature without reaching the fullness of the transitions between the basins.

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

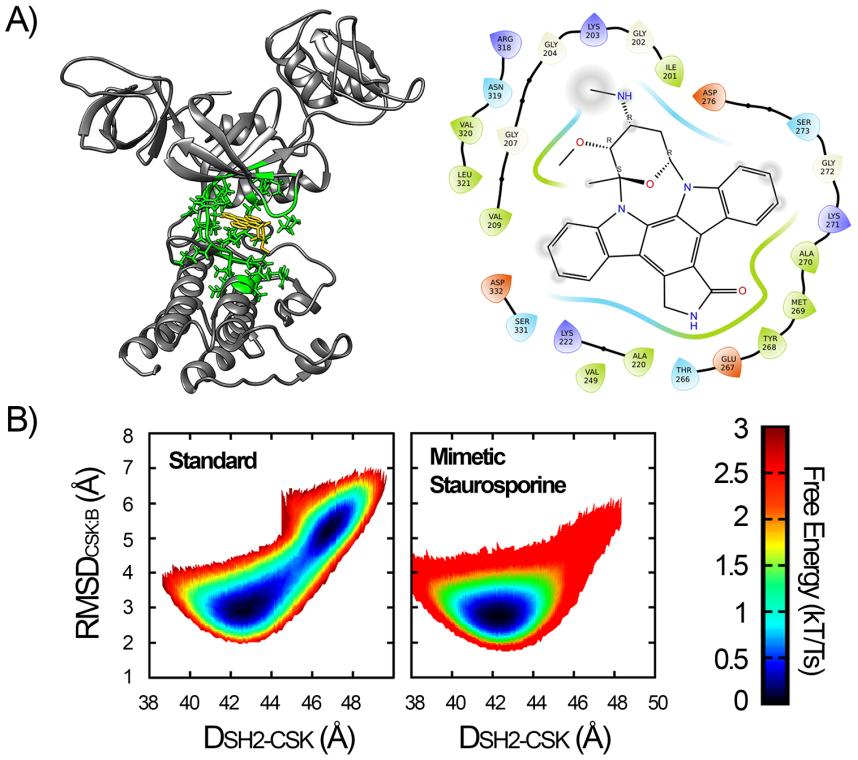


Figure 4. Effect of the inhibitor staurosporine through a mimetic potential under Cα-SBM simulations. (A) Three- (left) and two-dimensional (right) representation of staurosporine interactions with Csk. Residues highlighted in green in the cartoon are at a distance up to 6 Å from the ligand, presented in yellow. The 2D map of the interactions was generated using Maestro (Schrödinger Release 2022-2: Maestro, Schrödinger, LLC, New York, NY, 2021) where hydrophobic residues are in green, positively charged residues are in blue, negatively charged residues are in red, and apolar residues are in cyan. The gray atomic background represents the solvent-accessible surface area (SASA) of the atom. (B) Free energy surfaces for the Cα-SBM simulations with the dual potential without modifications (left) and mimetizing the staurosporine bound.

In summary, each of these highlighted dihedrals directly affects conformational transitions by increasing the energy barrier between the two states. Consequently, when all are stressed simultaneously, the transitions stop completely, making Csk stay predominantly in just one state during molecular dynamics simulations. Also, the obtained energetic profiles reinforce the relationship between global and local movements and the SH2 and SH3 modulation, indicating that the αC helix is a promising region for further studies, which also goes through conformational variations during the transition.

Effects of Domain Deletion for the Conformational Transitions. The importance of SH3 and SH2 coordination[23,48] and influence on the Csk activity[15] were explored by the removal of these domains. Computational simulations were performed through Cα-SBM molecular dynamics with the deletions of SH3 (ΔSH3) and SH3−SH2 (ΔSH3−ΔSH2) domains, as represented by cartoons in Figure 3A. Based on the free energy profiles shown in Figure 3B, using $D_{SH2-CSK}$, it is noted that the absence of the SH3 domain destabilizes the dynamic coupling between the domains and by consequence the correlated movement of SH2 and αC helix. Yet, SH2 has a displacement more restricted to a $D_{SH2-CSK}$ around 43 Å, where previously the typical values were 42 and 47 Å, relative to the closed and open forms, respectively. Such behavior, keeping the SH2 domain in a position closer to the closed conformation after the SH3 deletion, raises the speculation that the functional activity may be related to the overall movement of Csk and not limited to a specific state, which should be the subject of investigation in future works. Thus, deletion of the SH3 domain partially affected the conformational transition of Csk decreasing the number of conformational transitions.

The deletion of the SH3 and SH2 domains together vanishes the transitions, with $RMSD_{\alpha C\ helix}$ becoming uncoordinated with the well values around the closed form basin.

Effect of a Kinase Inhibitor for the Transitional Changes: A Starting Point for the Rational Design of New Drugs. Another interesting point is to understand how an inhibitor affects the mechanisms related to Csk conformational transitions by its proximity to the αC helix residues and restraints imposed by the molecular interactions. Thus, staurosporine (STU), a known inhibitor among Csk-like protein kinases,[40,50] was chosen for the evaluation, mainly due to the availability of a crystallographic structure (PDB ID: 3D7T). Initially, the results of Cα-SBM with the mimetic complexed potential (see Methods) were compared with the noncomplexed (standard) dual basin potential. Figure 4A shows Csk complexed with staurosporine (left) represented by a cartoon and the residues highlighted with sticks that were used to build the mimetic potential (right). Figure 4B shows the free energy surface using $D_{SH2-CSK}$ and $RMSD_{CSK:B}$ as reaction coordinates, indicating an ensemble limited to the closed form to the mimetic potential. Of course, the closed form is populated due to the way

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

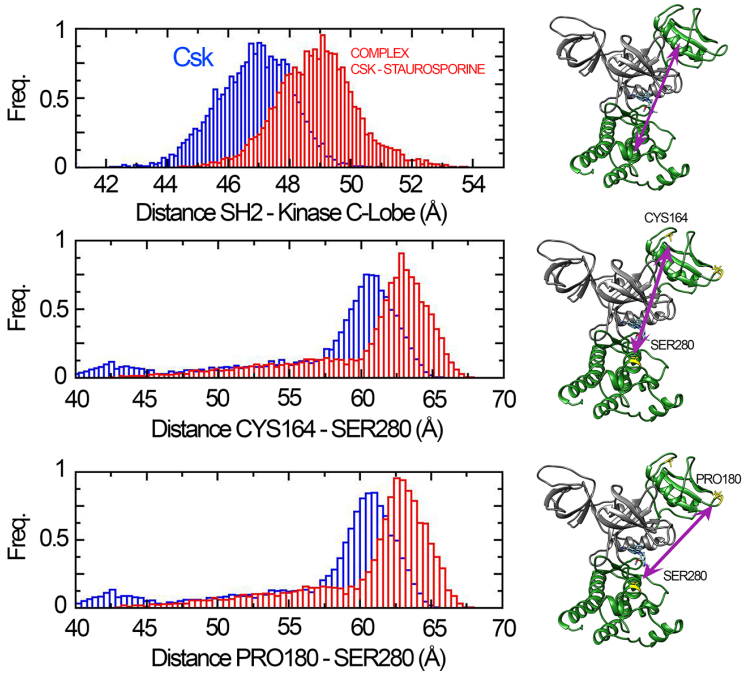


Figure 5. Detailed molecular dynamics simulations of Apo Csk and complexed with staurosporine. The histograms on the left show distributions of distance (distances between the center of mass of the SH2 domain and C-terminal Lobe Kinase, CYS164 and SER280, and PRO180 and SER280) to the Apo Csk (blue) and CSK−Staurosporine complex (red) obtained from simulations with a respective cartoon representation on the right. Purple double arrows highlight the distances involved with sparse residues in yellow.

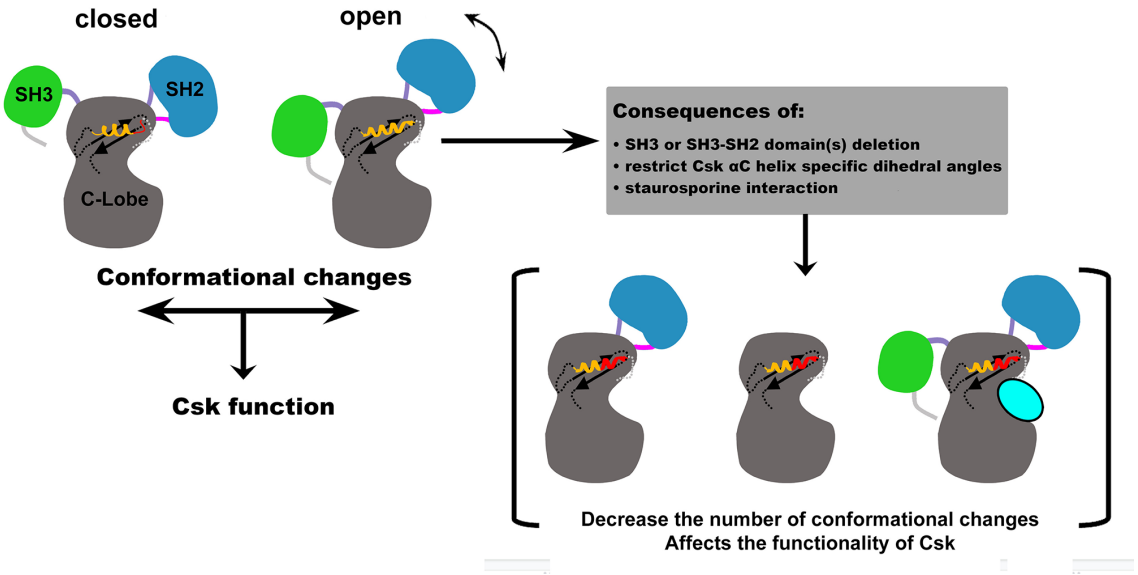


Figure 6. Summary of the elements that affect Csk conformations transitions directly related to the αC helix.

the dual potential is constructed. However, the mimetic potential practically vanishes the transitions between the two basins. Again, the understanding is that STU affects the transitions and as a consequence the Csk activity.

Since the closed form is understood as being active and its population obtained from Cα-SBM for the mimetic potential can be an artifact of the construction, detailed molecular dynamics simulations with explicit atoms using GROMACS were performed to evaluated Csk without any ligand and complexed with staurosporine. It is known that these molecular dynamics do not allow conformational changes as calculated using SBM; in contrast, they allow one to verify the influence of the ligand for structural stabilization of the complex, small movements including side chains and residue fluctuations.

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The detailed molecular dynamics simulations indicated a stable complex. The evaluation of the regime of residue fluctuations did not show significant variations for both cases (Figure S5). However, the analysis of the trajectories with and without the STU inhibitor revealed different distance distributions under different parameters. Assuming the histogram distributions in Figure 5, calculated using the distances between the centers of mass SH2 and C-terminal Kinase (top), CYS164-SER280 (middle), and PRO180-SER280 (bottom), note that Csk may have some structural movements which do not occur in the dynamics of the complex, based on the peaks despite the similar widths. The complex shows a profile with values of these chosen parameters, for example, the peaks to distances between the centers of mass SH2 and C-terminal Kinase being 47 and 49 Å to noncomplexed and complexed states, respectively. Here, it is evident that for both, Cα-SBM and detailed molecular dynamics simulations, the main effect of staurosporine in Csk in conformational terms is to prevent large movements of the SH2 domain, making the transitions between open and closed states difficult to occur in complexed Csk (holo) than in Apo. A summary of the elements that affect the Csk conformational transitions directly related to the αC helix region is presented in Figure 6.

DISCUSSION

Although having structural similarity to other SFK members, Csk differs in its structural arrangement and functioning mechanism.[11,40] For SFKs, the catalytic activity is negatively modulated by interactions between a phosphorylated C-terminal tail and the SH2 domain.[4] These interactions can be reversed by other proteins such as Cbp/PAG phosphatase that dephosphorylate the C-terminal tail and undo the interaction. Csk lacks this phosphorylated C-terminal tail and as consequence presents a regulatory mechanism different from the other SKFs. Thereby, Csk conformational changes are strongly related to the modulation between its domains, including specific regions. The results shown here indicate that αC helix is an essential region for these functional transitions since it establishes a coordination between the Csk domains and acts as a controlling region where local conformational changes are directly related to global conformational changes. This feature provides a considerable potential for the rational design of inhibitors with more competitive costs than STU, especially if the interaction mechanism is able to restrict the degree of freedom of the Csk αC helix. The importance of this region has already been reported by several authors;[47,49] however, the association of the restrictions in the αC helix as a result of the inhibitor interaction pushes forward the knowledge about how this small region works as a trigger for the Csk conformational change and by consequence on its activity. Still, it is important to highlight that the computational simulations employed do not evaluate the activity properly; i.e., this is only an indirect hypothesis associated with the conformational transitions or the possibility of the active (closed) form to be reached.

Following the same reasoning, the study of the role of an inhibitor (STU) by an additional layer of sophistication brought by the detailed molecular dynamics simulations supported the evidence pointed by the simple Cα-SBM mimetic construction and the stiffening of dihedral angles along the αC helix specific residues, in terms of making the transitions more difficult in the presence of STU. Still, when the detailed molecular dynamics simulations were explored, a higher distance from the SH2 domain and the kinase lobes was found, which in some way corroborates with the Jamros et al. SAXS characterization[12] where conformational shift effects were found in other Csk interactors. In the mentioned study, the Apo Csk, AMP-PNP (an analog of the ATP) bound Csk, and the ADP bound Csk showed different sets of 3 prominent conformational populations, varying the radius of gyration from lower values for Apo Csk to higher values for ADP bound Csk. This result corroborates with that presented in Figure 5 where the Apo Csk shows lower distances between the SH2 domain and the Csk C-Lobe center of mass than those in the Csk−Staurosporine complex, since in this case, the radius of gyration and SH2-Csk C-Lobe distance are correlated. The analysis also suggests that the interaction of different molecules with the Csk αC helix may be the main reason for allosteric regulation.

Still, the comparison of the Cα-SBM simulations employing the complete Csk and considering SH3 and SH2−SH3 domain deletions also corroborate with the experimental findings previously reported,[14,18,22,24] the activity values of 70% and 96%, respectively. Initially, it was possible to verify how important the coordination of SH2 and SH3 domains is with the αC helix and the Csk lobes, since the deletion results in a wilder representative ensemble without a clear population of two basins when evaluated under the $RMSD_{\alpha C\ helix:B}$ parameter, and related to a reduction of the Csk activity in experimental reports. These results can assume at least two hypotheses to relate computational simulations to the Csk activity, reserving the limitations of the modeling: (I) the Csk becomes less active due to staying a shorter time in the closed form, considered the active state, since the number of transitions decreases compared to the complete Csk; (II) the action of the Csk, as a molecular motor, occurs during the conformational transition, like scissors only cut when their blades are moving, in a naive comparison. The results discussed in this work allow even more speculations about how Csk conformational transitions in itself are related with movements where transitional states may be responsible for its regulation mechanism, these hypotheses being the subject of future work combining theoretical and experimental techniques.

CONCLUSIONS

The current work reinforces some previous results about the importance of the Csk αC helix region for the global Csk conformational transitions. By using Cα-SBM, a dual-basin potential was constructed allowing the evaluation of conformational transitions between the Csk open and closed forms, based on the X-ray crystallography (PDB ID: 1K9A:B and 1K9A:C, respectively). Also, the introduction of restrictions in some αC helix dihedral angles, deletions of SH3 and SH2−SH3 domains, and construction of an inhibitor mimetic potential indicated how these perturbations affect the conformational transitions, in some cases vanishing them. Detailed molecular simulations were employed as an additional technique to explore the relationship between Csk elongation degree as a consequence of staurosporine interaction, a known inhibitor. This analysis still opens possibilities for new approaches and investigations, like drug rational design based on the variations in Csk αC helix as an indicator of conformational changes. Finally, the results corroborate with experimental evaluations previously reported by different authors, expanding the knowledge involving Csk mechanisms.

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

The Journal of Physical Chemistry B pubs.acs.org/JPCB Article

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c05408.

Free energy profiles under a unique coordinate reaction to Cα-SBM dual potential, analysis of the dihedral angle torsions of the Csk αC helix during Cα-SBM simulations using dual potential, absence of conformational transitions using nonmodified Cα-SBM potentials, detailing of the Cα-SBM dual potential construction, and analysis of residue deviation, fluctuation in explicit water and ions all-atoms MD simulations, and free energy profiles using all-atoms SBM simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Leandro Cristante de Oliveira − *São Paulo State University (Unesp), Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo 15054-000, Brazil;* orcid.org/0000-0002-6932-6792; Phone: +55 (17)3221-2200; Email: leandro.cristante@unesp.br

Authors

Raphael Vinicius Rodrigues Dias − *São Paulo State University (Unesp), Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo 15054-000, Brazil*

Carolina Tatiani Alves Ferreira − *São Paulo State University (Unesp), Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo 15054-000, Brazil*

Patricia Ann Jennings − *University of California, San Diego, La Jolla, California 92093, United States;* orcid.org/0000-0002-7478-2223

Paul Charles Whitford − *Northeastern University, Department of Physics and Center for Theoretical Biological Physics, Boston, Massachusetts 02115, United States;* orcid.org/0000-0001-7104-2265

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c05408

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank Professor José Nelson Onuchic for the discussions about the Csk conformational transitions and support and Professor Fernando Alves de Melo for the discussions about the SH3 and SH2 domains and their role in protein kinases. R.V.R.D. would like to thank FAPESP and CAPES and C.T.A.F. would like to thank CAPES for the scholarships. This research was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico − CNPq 442352/2014−0 and Fundação de Amparo à Pesquisa do Estado de São Paulo − FAPESP grant no. 2016/08753-6. P.C.W. was supported by NSF grant MCB-1915843. Work at the Center for Theoretical Biological Physics was also supported by the NSF (Grant PHY-2019745). The simulations were performed at the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) and CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em Sao Paulo), project UNICAMP/FINEP-MCTII. Molecular graphics were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. We would like to thank the anonymous reviewers for their helpful comments.

■ REFERENCES

(1) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.* **2007**, *366*, 1661−71.

(2) Lopez, E. D.; Burastero, O.; Arcon, J. P.; Defelipe, L. A.; Ahn, N. G.; Marti, M. A.; Turjanski, A. G. Kinase Activation by Small Conformational Changes. *J. Chem. Inf. Model* **2020**, *60*, 821−832.

(3) Orellana, L. Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier. *Front. Mol. Biosci.* **2019**, *6*, 117.

(4) Okada, M. Regulation of the Src family kinases by Csk. *Int. J. Biol. Sci.* **2012**, *8*, 1385−1397.

(5) Thomas, S. M.; Brugge, J. S. Cellular functions regulated by SRC family kinases. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 513−609.

(6) Kawabuchi, M.; Satomi, Y.; Takao, T.; Shimonishi, Y.; Nada, S.; Nagal, K.; Tarakhovsky, A.; Okada, M. Transmembrane phosphoprotein Cbp regulates the activities of Src-family tyrosine kinases. *Nature* **2000**, *404*, 999−1003.

(7) Brdička, T.; et al. Phosphoprotein associated with glycosphingolipid-enriched microdomains (PAG), a novel ubiquitously expressed transmembrane adaptor protein, binds the protein tyrosine kinase Csk and is involved in regulation of T cell activation. *J. Exp. Med.* **2000**, *191*, 1591−1604.

(8) Nada, S.; Yagi, T.; Takeda, H.; Tokunaga, T.; Nakagawa, H.; Ikawa, Y.; Okada, M.; Aizawa, S. Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell* **1993**, *73*, 1125−1135.

(9) Talamonti, M. S.; Roh, M. S.; Curley, S. A.; Gallick, G. E. Increase in activity and level of pp60c-src in progressive stages of human colorectal cancer. *J. Clin. Investig.* **1993**, *91*, 53−60.

(10) Ogawa, A.; Takayama, Y.; Sakai, H.; Chong, K. T.; Takeuchi, S.; Nakagawa, A.; Nada, S.; Okada, M.; Tsukihara, T. Structure of the carboxyl-terminal Src kinase, Csk. *J. Biol. Chem.* **2002**, *277*, 14351−14354.

(11) Jamros, M. A.; Oliveira, L. C.; Whitford, P. C.; Onuchic, J. N.; Adams, J. A.; Blumenthal, D. K.; Jennings, P. A. Proteins at work: A combined small angle x-ray scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *J. Biol. Chem.* **2010**, *285*, 36121−36128.

(12) Jamros, M. A.; Oliveira, L. C.; Whitford, P. C.; Onuchic, J. N.; Adams, J. A.; Jennings, P. A. Substrate-Specific Reorganization of the Conformational Ensemble of CSK Implicates Novel Modes of Kinase Function. *PLoS Comput. Biol.* **2012**, *8*, e1002695.

(13) Lieser, S. A.; Shaffer, J.; Adams, J. A. Src tail phosphorylation is limited by structural changes in the regulatory tyrosine kinase Csk. *J. Biol. Chem.* **2006**, *281*, 38004−38012.

(14) Sun, G.; Budde, R. J. Mutations in the N-terminal regulatory region reduce the catalytic activity of Csk, but do not affect its recognition of Src. *Arch. Biochem.* **1999**, *367*, 167−172.

(15) Wong, L.; Lieser, S.; Chie-Leon, B.; Miyashita, O.; Aubol, B.; Shaffer, J.; Onuchic, J. N.; Jennings, P. A.; Woods, V. L.; Adams, J. A. Dynamic coupling between the SH2 domain and active site of the COOH terminal Src kinase, Csk. *J. Mol. Biol.* **2004**, *341*, 93−106.

(16) Shekhtman, A.; Ghose, R.; Wang, D.; Cole, P. A.; Cowburn, D. Novel mechanism of regulation of the non-receptor protein tyrosine kinase Csk: Insights from NMR mapping studies and site-directed mutagenesis. *J. Mol. Biol.* **2001**, *314*, 129−138.

(17) Cole, P. A.; Shen, K.; Qiao, Y.; Wang, D. Protein tyrosine kinases Src and Csk: A tail's tale. *Curr. Opin. Chem. Biol.* **2003**, *7*, 580−585.

(18) Lamers, M. B.; Antson, A. A.; Hubbard, R. E.; Scott, R. K.; Williams, D. H. Structure of the protein tyrosine kinase domain of C-

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

terminal Src kinase (CSK) in complex with staurosporine. *J. Mol. Biol.* **1999**, *285*, 713−725.
(19) Shaffer, J.; Sun, G.; Adams, J. A. Nucleotide release and associated conformational changes regulate function in the COOH-terminal Src kinase, Csk. *Biochem.* **2001**, *40*, 11149−11155.
(20) Songyang, Z.; Shoelson, S. E.; McGlade, J.; Olivier, P.; Pawson, T.; Bustelo, X. R.; Barbacid, M.; Sabe, H.; Hanafusa, H.; Yi, T. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol. Cell. Biol.* **1994**, *14*, 2777−2785.
(21) Ghose, R.; Shekhtman, A.; Goger, M. J.; Ji, H.; Cowburn, D. A novel, specific interaction involving the Csk SH3 domain and its natural ligand. *Nat. Struct. Biol.* **2001**, *8*, 998−1004.
(22) Sondhi, D.; Cole, P. A. Domain interactions in protein tyrosine kinase Csk. *Biochem* **1999**, *38*, 11147−11155.
(23) Barkho, S.; Pierce, L. C.; Li, S.; Adams, J. A.; Jennings, P. A. Theoretical insights reveal novel motions in Csk's SH3 domain that control kinase activation. *PLoS One* **2015**, *10*, e0127724.
(24) Lin, X.; Ayrapetov, M. K.; Lee, S.; Parang, K.; Sun, G. Probing the communication between the regulatory and catalytic domains of a protein tyrosine kinase, Csk. *Biochem.* **2005**, *44*, 1561−1567.
(25) Mills, J. E.; Whitford, P. C.; Shaffer, J.; Onuchic, J. N.; Adams, J. A.; Jennings, P. A. A novel disulfide bond in the SH2 Domain of the C-terminal Src kinase controls catalytic activity. *J. Mol. Biol.* **2007**, *365*, 1460−8.
(26) Chahine, J.; Cheung, M. S. Computational studies of the reversible domain swapping of p13suc1. *Biophys. J.* **2005**, *89*, 2693−2700.
(27) Gosavi, S.; Chavez, L. L.; Jennings, P. A.; Onuchic, J. N. Topological frustration and the folding of interleukin-1β. *J. Mol. Biol.* **2006**, *357*, 986−996.
(28) Oliveira, L. C.; Schug, A.; Onuchic, J. N. Geometrical features of the protein folding mechanism are a robust property of the energy landscape: A detailed investigation of several reduced models. *J. Phys. Chem. B* **2008**, *112*, 6131−6136.
(29) Noel, J. K.; Chahine, J.; Leite, V. B.; Whitford, P. C. Capturing Transition Paths and Transition States for Conformational Rearrangements in the Ribosome. *Biophys. J.* **2014**, *107*, 2881−2890.
(30) Cota, J.; Oliveira, L. C.; Damasio, A. R.L.; Citadini, A. P.; Hoffmam, Z. B.; Alvarez, T. M.; Codima, C. A.; Leite, V. B.P.; Pastore, G.; de Oliveira-Neto, M.; Murakami, M. T.; Ruller, R.; Squina, F. M. Assembling a xylanase-lichenase chimera through all-atom molecular dynamics simulations. *Biochim. Biophys. Acta* **2013**, *1834*, 1492−1500.
(31) de Oliveira, L. C.; da Silva, V. M.; Colussi, F.; Cabral, A. D.; de Oliveira Neto, M.; Squina, F. M.; Garcia, W. Conformational Changes in a Hyperthermostable Glycoside Hydrolase: Enzymatic Activity Is a Consequence of the Loop Dynamics and Protonation Balance. *PloS one* **2015**, *10*, No. e0118225.
(32) Clementi, C.; Nymeyer, H.; Onuchic, J. N. 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. *J. Mol. Biol.* **2000**, *298*, 937−953.
(33) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* **2012**, *116*, 8692−8702.
(34) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins: Struct. Funct. Genet.* **2009**, *75*, 430−441.
(35) Jackal is supported by funding from the National Science Foundation and National Institutes of Health Grant # DBI-9904841 and 5 R37 GM30518; http://honig.c2b2.columbia.edu/jackal.
(36) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12*, e1004794.
(37) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(38) Berendsen, H. J.; Postma, J. P.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *Chem. Phys.* **1984**, *81*, 3684−3690.
(39) Swendsen, R. H. Modern methods of analyzing Monte Carlo computer simulations. *Phys. A: Stat. Mech. Appl.* **1993**, *194*, 53−62.
(40) Levinson, N. M.; Seeliger, M. A.; Cole, P. A.; Kuriyan, J. Structural Basis for the Recognition of c-Src by Its Inactivator Csk. *Cell* **2008**, *134*, 124−134.
(41) Russell, R. B.; Barton, G. J. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins: Struct. Funct. Genet.* **1992**, *14*, 309−323.
(42) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(43) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359−2368.
(44) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(45) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(46) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(47) Huang, H.; Zhao, R.; Dickson, B. M.; Skeel, R. D.; Post, C. B. αC helix as a switch in the conformational transition of Src/CDK-like kinase domains. *J. Phys. Chem. B* **2012**, *116*, 4465−4475.
(48) Wong, L.; Lieser, S. A.; Miyashita, O.; Miller, M.; Tasken, K.; Onuchic, J. N.; Adams, J. A.; Woods, V. L.; Jennings, P. A. Coupled motions in the SH2 and kinase domains of Csk control Src phosphorylation. *J. Mol. Biol.* **2005**, *351*, 131−143.
(49) Onuchic, J. N.; Kobayashi, C.; Miyashita, O.; Jennings, P.; Baldridge, K. K. Exploring biomolecular machines: Energy landscape control of biological reactions. *Philos. Trans. R. Soc. London B: Biol. Sci.* **2006**, *361*, 1439−1443.
(50) Kinoshita, T.; Matsubara, M.; Ishiguro, H.; Okita, K.; Tada, T. Structure of human Fyn kinase domain complexed with staurosporine. *Biochem. Biophys. Res. Commun.* **2006**, *346*, 840−844.

https://doi.org/10.1021/acs.jpcb.2c05408
J. Phys. Chem. B XXXX, XXX, XXX−XXX

B Trabalhos Submetidos e em preparação.

Article

Unveiling the GRB2 folding mechanisms through the characterization of metastable ensembles and pathways

Raphael Vinícius Rodrigues Dias [1], Renan Pereira Pedro [1,2], Murilo Nogueira Sanches [1], Vitor Barbanti Pereira Leite [1], Ícaro Putinhon Caruso [1,2,3], Fernando Alves de Melo [1,2,*] and Leandro Cristante de Oliveira [1,*]

[1] Department of Physics, Institute of Biosciences, Humanities and Exact Sciences (IBILCE), São Paulo State University "Júlio de Mesquita Filho" (UNESP), São José do Rio Preto, SP, Brazil.
[2] Multiuser Center for Biomolecular Innovation (CMIB), São Paulo State University "Júlio de Mesquita Filho" (UNESP), São José do Rio Preto, SP, Brazil.
[3] Institute of Medical Biochemistry Leopoldo de Meis (IBqM) and National Center for Structural Biology and Bioimaging (CENABIO), Federal University of Rio de Janeiro (UFRJ), 21941-590 Rio de Janeiro, RJ, Brazil
* Correspondence: fernando.melo@unesp.br (F.A.M.); leandro.cristante@unesp.br (L.C.O.)

Abstract: The growth factor receptor-bound protein 2 (GRB2) adaptor protein is essential in regulating cell survival. Its structure consists of one central Src Homology 2 (SH2) domain flanked by two Src Homology 3 (SH3) domains. While SH2 is essential for interacting with phosphotyrosine regions in the other proteins, SH3 is responsible for recognizing rich proline regions on protein partners during cell signaling. In this study, we combined computational and experimental approaches to investigate the folding mechanism of GRB2. Through molecular dynamics simulations that sample the phase space, and the Energy Landscape Visualization Method (ELViM), we were able to visualize a 3-dimensional representation of the globally funneled energy surface. Consequently, we identify two possible parallel folding routes that can not be seen in a 1-dimensional analysis. Finally, the experimental results agreed with the literature, confirming preliminary mechanisms supporting the hypotheses proposed in this study.

Keywords: keyword 1; Structure based simulations; Domain folding; Folding funnel; thermodynamics stability.

Citation: Dias, R.V.R.; Pedro, R.P.; Sanches, M.N. Unveiling the GRB2 folding mechanisms through the characterization of metastable ensembles and pathways. *Journal Not Specified* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2023 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Growth-factor receptor-bound protein 2 (GRB2) is a 25 kDa adaptor protein comprised of a central SH2 domain (Src Homology 2) flanked between an N- and C-terminal SH3 domain (Src Homology 3) in the overall modular architecture of SH3N–SH2–SH3C, in a total of 217 residues (Figure 1A) [1].

The SH2 domains are present in more than 100 different proteins. They are found mainly in multicellular organisms, being part of protein kinases, transcription factors, cytokine signaling suppressors, and adapter proteins, thus indicating the importance of the domain in intracellular communication [2–5]. This domain is about 100 amino acids long that forms a globular module with a well-conserved secondary structure composed of a central β-strands antiparallel flanked by two α-helix on each side. It is remarkable to mention that there are two specific interaction sites (Site I and Site II), which are mechanistically crucial to how this domain mediates the protein-protein interaction in cell signaling pathways [6–9]. Site I consist of a pocket containing three positively charged amino acid residues (GRB2-SH2: ARG67, ARG86, and LYS109), whose side chains mediate cell signaling phosphotyrosine (pY) recognition. Site II is a hydrophobic pocket located nearby α_3-helix responsible for recognizing pY+1 to pY+3 peptides on protein partners [10–13].

The SH3 domains are protein modules found in more than 220 different proteins, performing a pivotal role within the cell, such as regulation of cellular pathways, cell proliferation, migration, and cytoskeleton modification [14–17]. These domains have a

secondary structure formed by a small 3_{10}-helix and also from four to eight β-strands in an arrangement of two β-sheets or β-barrel; in addition, the domain handles are named RT, n-Src and distal, being essential regions in recognition of partners for interaction [16,18,19]. Other studies have initially revealed that the SH3 domain has an affinity for interaction in PxxP motifs, accommodating it in two different sites [20,21]. Furthermore, peptides containing the PxxP motif could interact through two opposite conformations, defined by the relative position of a positively charged residue (+xxPxxP or xPxxPx+). Such conformations are essential in how the domain interacts with other proteins [22,23].

The GRB2 crystal structure (PDB ID: 1GRI.) shows that the protein crystallizes as a dimer where both protomers fit symmetrically in a way that the SH2 domain from one protomer faces the C-terminal SH3 domain from the other protomer and vice-versa. This structural arrangement is strategically self-assembled to stabilize the dimer while the SH2 and SH3 binding sites are ready to interact with other protein partners and ligands. GRB2 monomer/dimer equilibrium changes confer different action mechanisms, such as increasing protein expression level, phosphorylation, mutations, and interactions with ligands [24]. These factors influence, for example, cases where GRB2 overexpression occurs, significantly increasing the dimer population and suppressing MAPK activation. In contrast, dimer dissociation occurs by the upregulation of receptor tyrosine kinase (FGFR2) activity, where the overexpression of Y160 phosphorylated GRB2 monomer, leads to an increase in MAPK activation [25]. To summarize, GRB2 works as a switch-like mechanism where only its monomeric state is signaling competent to activate MAPK downstream, whereas dimeric GRB2 is the inhibitory form.

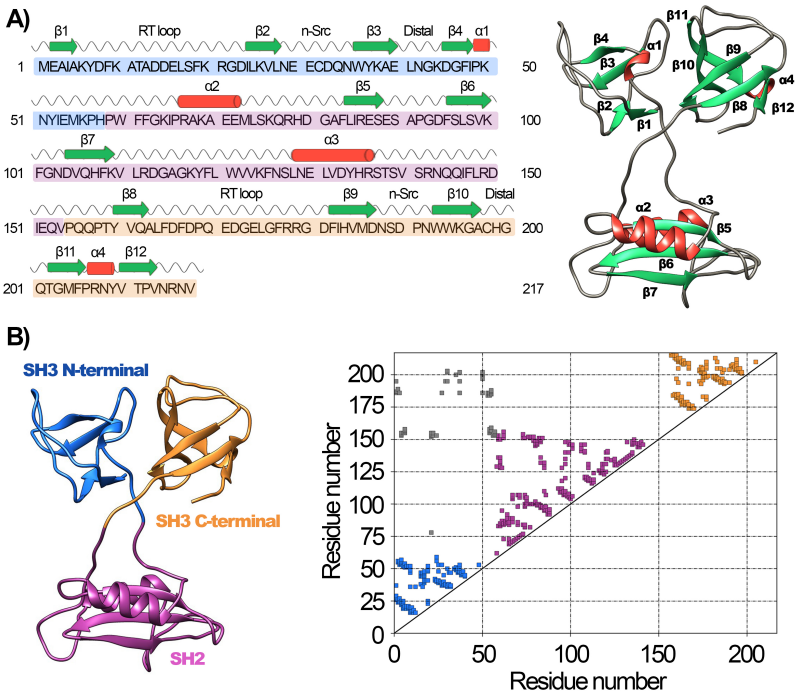


Figure 1. Structure of Monomeric GRB2 protein: (A) Sequence of GRB2m protein with the domains highlighted by colors: GRB2-SH3N blue, GRB2-SH2 purple, and GRB2-SH3C orange. Loops, coils, and secondary structures are also shown, with β-sheets highlighted in green arrows and α-helices in red cylinders. A right-side 3-D view shows these structures. (B) GRB2 structure shows the domains highlighted by colors as presented in (A), and the contact map calculated by the CSU method [26] shows all contacts of the native structure of GRB2.

Despite all these findings, studies that list the action mechanism of the GRB2 monomer with its structural flexibility and intermediate folding states are still scarce. Then, this study used computational and experimental techniques to assess structural changes during the folding of the GRB2 monomer (GRB2m), possible intermediate states, domain flexibility and conformational stability in order to elucidating the physical and structural mechanisms of GRB2 monomer. Also, this study can facilitate insights upon experiments, such as ligand construction and domain deletion, that can associate folding intermediate states with GRB2m action mechanism in cell.

2. Materials and Methods

2.1. Cα Structure-Based Model

The Structure-Based Model (SBM) has been efficient in investigating kinetic and thermodynamic properties in protein folding processes since its potential is defined in such a way that the minimum energy is attributed to the native state of the protein [27,28]. In addition to this simplification associated with the potential for interaction between the components of the system, the components themselves are also simplified, using the Cα model, in which amino acids are represented only by their alpha carbons, which are treated as spheres of the same mass and radius. The SBM potential is given by a harmonic term for the bonds and another for the angles. The contact term is a Lennard-Jones potential, with 10-12 instead of 6-12 for better sampling the minimum energy [28], and the non-contact term takes into account repulsion interactions between pairs of residues that are not in contact in the native structure. In this model, the potential of a Γ configuration of a protein which has a configuration in the native Γ_o state, is expressed by:

$$\begin{aligned} V(\Gamma,\Gamma_o) &= \sum_{bonds} \epsilon_r (r-r_o)^2 + \sum_{angles} \epsilon_\theta (\theta-\theta_o)^2 \\ &+ \sum_{backbone} \epsilon_\phi \left\{ [1-\cos(\phi-\phi_o)] + \frac{1}{2}[1-\cos\left(3(\phi-\phi_o)\right)] \right\} \\ &+ \sum_{contacts} \epsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \end{aligned} \quad (1)$$

where r_0, θ_0 e ϕ_0 defined by the distance between two subsequent residues, angles formed by three subsequent residues and dihedrals formed by four subsequent residues, respectively. d_{ij} is the native distance between contact pair i and j determined by CSU software (Contact of Structural Units) [26], ϵ_C = 1.0 kcal/mol, ϵ_r = 100 kcal/(mol Å), ϵ_θ = 20 kcal/mol, ϵ_{NC} = 1.0 kcal/mol and σ_{NC} = 4.00 Å . Structure-based models are well detailed in several references [27–30], and widely used in folding and protein mechanisms studies.

2.2. Simulation Details and thermodynamics analysis

The input files for structure-based simulations were generated by SMOG web server [31]. Molecular dynamics were performed using GROMACS 4.5.7 [32,33] where a temperature bath was coupled by Langevin Dynamics with a constant equal to 1 ps [30]. To estimate the folding temperature T_f, temperature at which 50% of the simulation time the protein is unfolded and 50% is folded, we performed constant temperature simulations. Simulations have a total of $5x10^8$ steps for each temperature evaluated and using a timestep of 0.0005 time units. We then performed a larger constant temperature simulation ($5x10^9$ steps) at the estimated T_f. The configurations and energies were recorded every 5000 steps. Different reaction coordinates were used to follow the folding and unfolding: Native Contacts (Q), q_w (Described in the ELViM method) and RMSD of the complete protein using the native structure (PDB ID: 1GRI chain A). All thermodynamic analysis involving statistical physics was generated using the WHAM method [34].

2.3. Energy Landscape Visualization Method (ELViM)

The protein folding process is usually analyzed using a reaction coordinate associated with the native state. However, it is not always possible to understand the energy surface of the system through only these coordinates. To visualize this process, an alternative is to employ the Energy Landscape Visualization Method (ELViM), a new method that uses a metric based on the internal distance between amino acids to describe the differences between two conformations [35]. This metric is described as:

$$q_w^{k,l} = \frac{1}{N_p} \sum_i \sum_j \exp[\frac{-(r_{i,j}^k - r_{i,j}^l)^2}{\sigma_{i,j}^2}] \quad (2)$$

where N_p is the total numbers of pairs of residues, $r^k(l)_{i,j}$ is the distance between the residue i and j in the conformation $k(l)$ and $\sigma_{i,j}$ is a weighing constant equals to $\sigma_0|i-j|^\epsilon$, with σ_0=1 and ϵ=0.15. We can then describe the dissimilarity of our phase space as $\delta^{k,l} = 1 - q^{k,l}$.

To visualize the energy surface we first compute the dissimilarity for all pairs of structures k and l of a given trajectory, resulting in a symmetric matrix. After that, a multidimensional reduction of the original data takes place, to reduce the original multidimensional space of the data to a dimensional space comprehensible to human perception, that is, a visual space of at most three dimensions. In this work we used the Force Scheme technique to reduce the multidimensional data [36]. This technique initially places the objects of the matrix at random initial positions in a 2D space and then applies a minimization step to preserve the distances in the dissimilarity matrix. The process stops when the difference in distances between consecutive iterations is below a given cut-off. The data analyzed by ELViM within this study consist of the conformations obtained at the folding temperature of GRB2m, to ensure that we could visualize and identify the routes and transition states between the free energy barriers.

2.4. Protein preparation

The construction of the pET-28a(+) vector for GRB2m was based on the structure of the full-length GRB2 [1], generally the GRB2 protein is expressed in the dimeric form and to express it in the monomeric form we perform a Y160F mutation. A 6xHis tag was added to the N-terminal portion of the protein to aid in affinity chromatography purification. GRB2m was expressed in *Escherichia.coli* BL21 (DE3) grown in Luria-Bertani (LB) broth and purified on IMAC HiTrap™ HP cobalt affinity column (GE Life Sciences). Further purification of the protein was performed on a Superdex® 75 size-exclusion column (GE Life Sciences) coupled to the GE Äkta FPLC™ system, where the protein was passed to a buffer of 20 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0) with 50 mM NaCl and 0.5 mM Phenylmethylsulfonyl fluoride (PMSF). Sample purity was analyzed by 15% SDS-PAGE. The concentration of the sample was evaluated through the absorbance at 280 nm using the molar extinction coefficient of 38,055 $M^{-1}\ cm^{-1}$. The molar extinction coefficient was obtained from the sequence of amino acid residues using the ProtParam online software on the ExPasy Server [37].

2.4.1. Circular Dichroism Spectroscopy

Thermally induced protein unfolding was evaluated by far-UV CD spectroscopy on a Jasco J-815 spectropolarimeter. The concentration of the protein used was 20 μM in 300 μl of solution, in a buffer of 20 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0) with 50 mM of NaF. It was used a quartz cuvette with an optical path of 1 mm placed in a sample holder controlled by Peltier Jasco CDF-426S. Spectra were collected in an acquisition range from 195 to 250 nm at 0,5 nm intervals. Measurements were taken at temperatures from 25 to 40 °C with increment of 5 °C and, after that, measurements were taken from 40 to 51 °C at 1 °C intervals. All measurements were performed three times. Protein secondary structure analyzes were performed using the CDPro software [38] using the CDSSTR [39] method.

2.5. Fluorescence Spectroscopy

Steady-state fluorescence spectroscopy measurements were performed on an ISS PC1 spectrofluorometer equipped with a quartz cuvette with a 1.0 cm optical path and Neslab RTE-111 refrigerated bath. The excitation wavelength was set at 295 nm and was chosen because it is the region where the tryptophan residue of the protein is excited. The emission spectrum was collected in the region between 305 to 500 nm, which was corrected for buffer background fluorescence [40]. The experiment was carried out by varying the temperature, starting from 10 to 80 °C with intervals of 5 °C, except in the region between 45 and 61 °C, where the measurement interval was 1 °C. The protein concentration was set at 5 μM in 2 ml of solution in a buffer of 20 mM Na_2HPO_4/NaH_22PO_4 (pH 7.0) with 50 mM of NaCl. In addition to the emission spectra, the center of spectral mass (CM) was calculated using the equation 3 [41,42]:

$$CM = \frac{\sum_i \lambda_i F_i}{\sum_i F_i} \tag{3}$$

where F_i is the emission in the wavelength λ_i.

2.6. Differential Scanning Calorimetry Thermal Analysis

DSC measurements were performed using a CSC 6300 Nano DSC III. Data were collected by heating the sample from 10 to 80 °C at a rate of 0.5 °C min^{-1}. The protein concentration was fixed at 80 μM in 700 μl of solution in a buffer of 20 mM Na_2HPO_4/NaH_22PO_4 (pH 7.0) with 50 mM of NaCl.

3. Results

3.1. Thermodynamic profile of GRB2m folding

We perform the simulations using a spectrum of temperatures so that the folded and unfolded ensembles are well defined. Temperatures around the folding temperature are widely simulated, so that multiple transitions occur between the equally populated folded and unfolded ensembles and the best sampling of the transition region is achieved. Using the Native Contacts (Q_{GRB2}) as the reaction coordinate, we calculated the free energy profile of GRB2m (Figure 2A). The result shows several intermediate states between the native and unfolded ensembles, meaning that the multiple domains fold into different folding temperatures (Tf), generating partially folded states. In this step, we clustered the structures referring to the minimum energy. Only one major conformation was expected in each cluster, however, in each of these there were overlaps of partially-folded states, possibly indicating several folding routes (Figure S1). Using the average of contacts formation in each domain (Figure 2B) and the specificity in the formation of contacts (Figure 2C), it was expected that in each well presented in the free energy profile there would be a specific structuring of a domain, such as the formation of both GRB2m-SH3 and then the formation of GRB2m-SH2 or vice versa. Due to limitation of the reaction coordinates used (Q_{GRB2}, RMSD and Radius of gyration) we lost overlap information of these values at different stages of folding, for example, in this case, the GRB2m-SH3C and GRB2m-SH3N domains of GRB2m present practically the same values of these parameters and the sum of these overlap with the values of GRB2m-SH2. Our hypotheses from these results would be a possible backtracking, requiring the formation and breaking of some structure for final formation, or that the folding of GRB2m should not be continuous in a specific sequence of folds of its domains, although there should be a preferential formation.

3.2. Funil landscape and partially-folded states.

Using the computational tool ELViM [35], we show that the folding of GRB2m does not always follow the specific continuous construction of the domains. In this representation, each dot corresponds to a conformation sampled by the molecular dynamics, and is colored according to the probability of contacts, with the z-axis corresponding to Q_{GRB2}. It's worth noting that the 3D projection (Figure 3) has a funnel-like shape, with the unfolded

Version January 10, 2023 submitted to *Journal Not Specified*

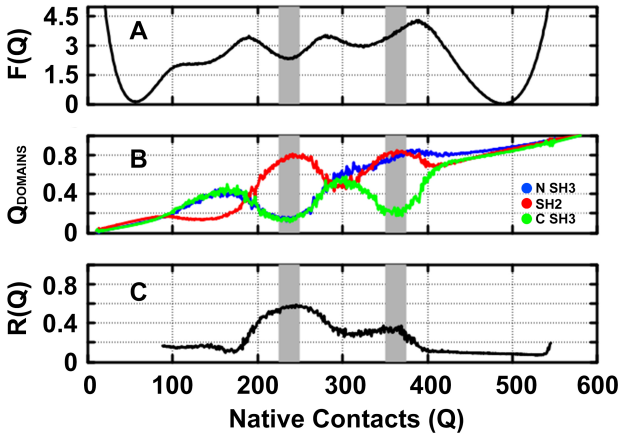


Figure 2. Folding details: Free Energy, Contacts and Route. A) Free Energy of GRB2m calculated from Cα-SBM simulations using Native Contacts of reaction coordinates, The Free Energy values are normalized by the transition state temperature (TS). B) Percentage of contacts formed along the reaction coordinate, each color represents the formation of contacts within a specific domain: blue - GRB2m-SH3 N-terminal, red - GRB2m-SH2 and green - GRB2m-SH3 C-terminal. Values are averages calculated over all dynamic frames at transition temperature. C) Folding route measurement R(Q) as a function of native contacts Q for the GRB2m dynamics on transition temperature.

conformations populating the upper region (dark blue dots) and the folded structures at the bottom (red dots). From this result we highlighted 6 regions, each one corresponding to a different combination of structured domains: (1) corresponding to the fully unfolded conformations; (2) corresponding to the partially-folded SH3N; (3) corresponding to both GRB2m-SH3 fully folded; (4) corresponding to the GRB2m-SH2 fully folded; (5) corresponding to the GRB2m-SH2 fully folded and the GRB2m-SH3N partially-folded, and (6) corresponding to the native conformations. Combining the funnel of structures ensembles with the calculation of the probability of contact formation, we distinguish two possible parallel folding routes in the structural formation of GRB2m domains: SH2-SH3N-SH3C and SH3C-SH3N-SH2 (Figure 3). In this new perspective, we discard the hypothesis of the occurrence of backtracking in the structural formation of GRB2m. It is worth mentioning that our calculations take into account the construction of tertiary structures of the GRB2m domains, not applying to possible local phenomena in secondary structures within each domain treated separately.

Since the calculated results use structural information taken from the dynamics, with equally spaced frames for maximum conformational space scanning, we decided to do a second analysis using consecutive frames to accurately capture the path during an unfolding/folding transition. Complementarily, we used the domain-separated free energy profile to verify the minimum energy paths, as well as the two-dimensional projection of conformational phase space, shown in Figure 4. We combined the methods and confirmed that both of them show two possibilities in the structural construction of GRB2m (paths). The preferred route, denoted as route 1, mostly accessed, consists of the formation of the GRB2m-SH2 domain, followed by the formation of GRB2m-SH3N domain and ending with the formation of GRB2m-SH3C domain, while the second route consists of the formation of GRB2m-SH3s domains together and finally the formation of GRB2m-SH2.

3.3. Domain deletion and thermodynamic stability mechanism.

We used the SBM model to build two new structures by deleting the GRB2m-SH3N and GRB2m-SH3C domains. Recently, Saki Ochi *et. al.* [43] showed experimentally that the deletion of these domains interfered with the stability of GRB2m. Among the results,

Version January 10, 2023 submitted to *Journal Not Specified*

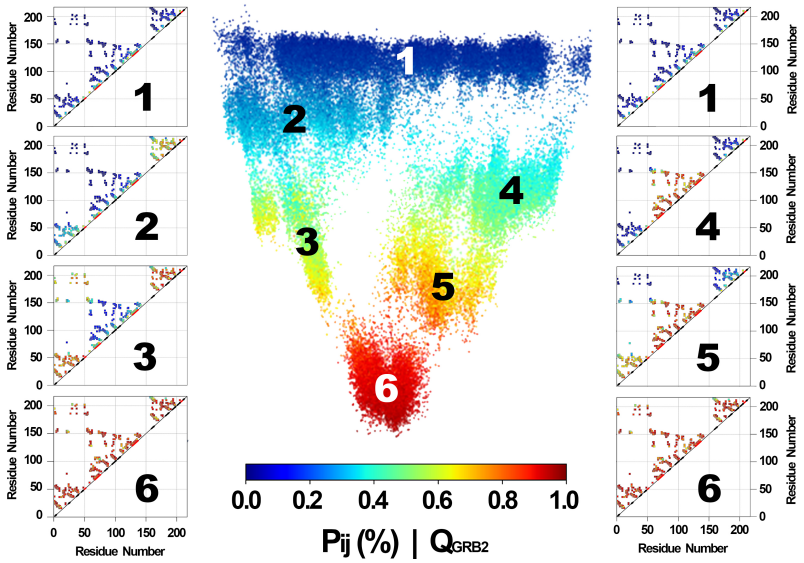


Figure 3. Contact probabilities and Funnil Landscape: Structural correlation between the energy funnel and the contact maps, highlighting the ensembles with each corresponding map. The lateral maps were colored according to the probability of contact formation (Pij). going from few formed contacts and unfolded domains (blue) to formed contacts and folded domains (red). The points of each conformation in funil representation are colored by the normalized fraction of native contacts (Q_{GRB2}).

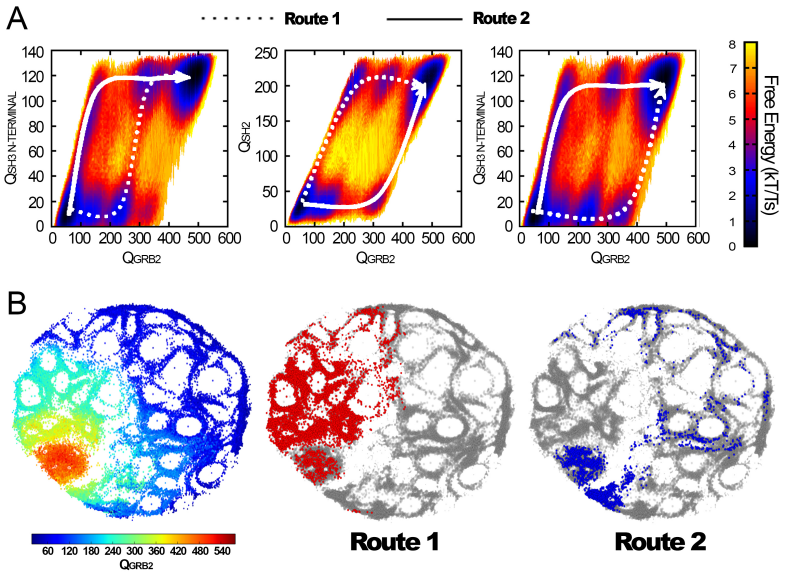


Figure 4. Minimal Energy Paths: A) Free Energy surface of GRB2m domains calculated from C-SBM simulations using as reaction coordinates the Native Contacts of using as reaction coordinates the native contacts of each specific domain. The Free Energy values follow the colorbar scale presented below from 0 (black) to 8 (yellow) normalized by the transition state temperature (Ts). B) Conformational phase space of GRB2m visualized through ELViM as a function of Q_{GRB2} and with the different routes highlighted in red and blue.

Version January 10, 2023 submitted to *Journal Not Specified*

they showed the importance of the N-terminal GRB2m-SH3 domain for the formation of the other domains, decreasing the transition temperature when deleted. In our results, we reproduce computationally the same thermodynamic effects, showing that the lack of the SH3N domain decreases the accessibility of the route 1 (SH2 – SH3C), spending a lot of time trapped in the ensemble with only the GRB2m-SH3C formed. As for the deletion of the GRB2m-SH3C domain, the transition ensembles are the same as for the complete GRB2m protein, with both routes well established. This set of information establishes the importance of the presence of the GRB2m-SH2 domain for the formation of GRB2m, showing that its contacts are a crucial parameter and that it establishes connections between the three domains.

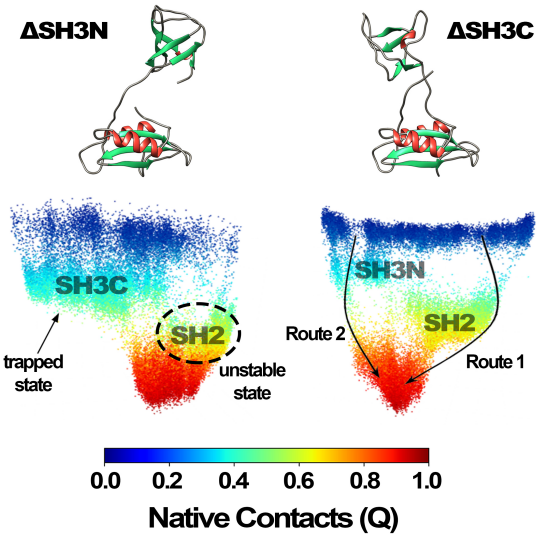


Figure 5. Domain deletions affect the paths of GRB2m folding. Structures used for molecular dynamics, colored according to secondary structure, with green representing β-sheets and orange α-helices. The energy funnels show that the deletion of N-terminal GRB2m-SH3 changes the folding profile of GRB2m, getting trapped in the ensemble with only C-terminal GRB2m-SH3 formed, yet the lack of this domain destabilized the formation of the SH2 domain. Deletion of the C-terminal GRB2m-SH3 does not affect the routes or destabilize the formation of domains.

3.4. DSC, Fluorescence and CD analysis

Due to the limitation of reduced computational models, we used in vitro experiments for qualitative comparison of results. Fluorescence spectroscopy and Differential Scanning Calorimetry (DSC) were used to check the structure of GRB2m with increase of temperature (Figure 6). The presence of tryptophans allowed the construction of the emission spectrum (Figure S2) and the calculation of the center of spectral mass (Figure 6). The results show unfolding curves characteristics with increasing temperature, with a Tf of approximately 51°C. This temperature was also obtained using the Spectral Center of Mass.

Although our experiments were carried out with the monomeric protein, the DSC profiles and the temperature found are in agreement with Saki Ochi et. al (43), who theorizes that the data of their results refer only to the SH2 domain of GRB2m. This hypothesis is supported by the fact that GRB2m-SH3s are rich in β-strands sensitive to thermal variations, exposing hydrophobic residues nearby the region, facilitating the creation of intermolecular hydrogen bonds and forming aggregates during the measurement [44,45].

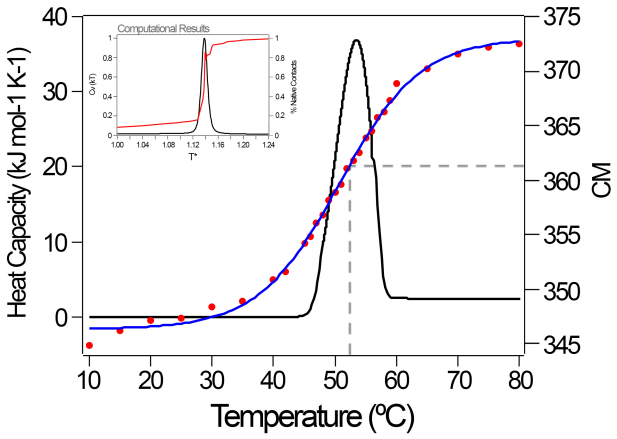


Figure 6. Thermodynamic profile of the state transition: Experimental data obtained through DSC and fluorescence, which presents a model of only two states (folded and unfolded) with a Tf of approximately 52 °C. Computational data similar to the experimental ones, corroborating the proposed model.

To confirm this, we perform the Circular Dichroism spectroscopy experiment (S. I. 3). Calculating the percentages of secondary structures along the temperatures, we verified an increase in the percentage of β-sheet, which was expected since the aggregation of the system tends towards this type of structure. On the other hand, a decrease in the α-helix was found and, knowing that this information can only come from the SH2 domain (Figure 1), this result indicates that all in-vitro folding analyzes should refer to this domain. Furthermore, the computational results show different folding routes, with close transition temperatures of domains, allowing the formation of partially-folded states simultaneously, i.e., at the same time that the temperature influences the GRB2m-SH3 of the protein in solution, it may also affect the GRB2m-SH2 domain, interfering and overlapping the signals of the obtained spectra.

4. Discussion

One of the fundamental and universal examples of biological self-assembly is the folding of proteins into their compact three-dimensional structures. In addition to its importance for biological activity, we also know that folding is coupled to many other biological processes, including the trafficking of molecules to specific cell sites and the regulation of cell growth and differentiation [45]. Understanding this complex process provides a unique view of the thermodynamic point, and from this understanding, it is possible to initiate several improvements, such as thermostability and affinity with molecules. The study of folding for small proteins and specific domains is already well established, however, there are few studies when it comes to folding proteins with multiple domains. In the last few years, some work has conflicted with the standard conventional assumption that each protein domain is an autonomous folding unit [45,47].

Here, we study the GRB2m protein, a multidomain protein formed by an GRB2m-SH2 domain flanked by two GRB2m-SH3s domains [1]. This is an important adaptor protein which is ubiquitously expressed in eukaryotic cells. It is involved in many intracellular protein interactions and readily detectable in malignant forms of colon, breast and prostate cancers for example [48,49]. Recently, Zamal et al [25] showed the importance of monomer-dimer balance for the function of GRB2m, where the monomeric form is the signaling-competent active form while the dimeric GRB2m is the inactive inhibitory form. Thus, we used the monomeric form of GRB2m to evaluate the folding of its domains, evaluating the

possible intermediate states, as well as the intercommunication and importance of these for its complete structural formation.

Our results show that GRB2m has a lot of flexibility due to the arrangement of its domains. We were also able to evaluate that during its folding process the protein can access partially-folded states, where one of its domains is in unfolded form while the others are folded. In addition, through our methodology, we built the free energy funnel that showed the different possibilities of folding, changing the folding sequences in relation to the GRB2m domains. This information is relevant since each domain has its well-defined characteristics and functions, for example, while GRB2m-SH3 is a recognition domain for other proteins that have proline-rich sequences, in the other hand, GRB2m-SH2 interacts with other molecules and phosphorylated proteins, being responsible even for the oligomerization of GRB2m [2–5,7–9,17–20,50,51]. All this complexity in the arrangement and function of its domains highlights the importance of different routes and different partially-folded states, which can adapt to the environment according to the needs imposed for the proper functioning of signaling pathways.

Following our model, we tested the results of Saki Ochi *et. al.* [43], where a thermal destabilization of GRB2m with the exclusion of the N-terminal GRB2m-SH3 domain is reported. We showed that the lack of the N-terminal GRB2m-SH3 domain interferes with the formation of the GRB2m-SH2 domain, trapping the protein and destabilizing its routes. In contrast, the C-terminal GRB2m-SH3 deletion does not interfere with the formation of the other domains. Our theoretical computational results corroborate the experiments in the literature, explaining that the interdomain contacts from the N-terminal SH3 domain are responsible for the structural stabilization of SH2 in the GRB2m monomeric structure. Also, the lack of C-terminal SH3 did not affect our thermodynamic profile, concluding that there is no structural need for this domain for the folding of the other two.

Finally, we performed in vitro experiments to complement our theoretical results, even if only by qualitative comparisons. Among the experiments, DSC and Fluorescence show that the folding temperature is approximately 52ºC, and this temperature is attributed to the folding of the GRB2m-SH2 domain, since we lost information due to structural limitations, as the fact that GRB2m-SH3s are rich in beta sheets and has a tendency to aggregate. To confirm this, CD experiments were performed at various temperatures, showing the decay of alpha helices and increasing beta sheet formation, as expected.

Author Contributions: R.V.R.D. and M.N.S., performed computer studies. R.P.P., performed experimental studies. V.B.P.L., I.P.C., F.A.M. and L.C.O., made the project idea and supervised the experimental work. R.V.R.D., R.P.P., M.N.S., V.B.P.L, V.B.P.L., I.P.C., F.A.M. and L.C.O., wrote the manuscript. All authors contributed to the analysis, discussion and critical feedback.

Acknowledgments: RVR was supported by FAPESP (grant) and CAPES (grant). MNS was supported by CAPES (grant 88887.661619/2022-00) and FAPESP (grant 2021/15028-4). RPP was supported by CNPq (grant 140306/2020-0). This research was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq 442352/2014–0 and Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (FERNANDO - 19/24974-0), FAPESP/CNPQ Vitor. The simulations were performed at the Center for Scientific Computing (NCC/Grid UNESP) of UNESP and CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em Sao Paulo), project UNICAMP/FINEP-MCTII. Molecular graphics performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper as potential competing interests.

Sample Availability: Samples of the compounds ... are available from the authors.

Abbreviations

The following abbreviations are used in this manuscript:

CM	Center of Spectral Mass
Tf	Folding Temperature
GRB2	Growth Factor Receptor-Bound Protein 2
GRB2m	GRB2 monomer
SH2	Src Homology 2
SH3	Src Homology 3
ELViM	Energy Landscape Visualization Method
SBM	Structure-Based Model
CSU	Contact of Structural Units

References

1. Maignan S, Guilloteau JP, Fromage N, Arnoux B, Becquart J, Ducruix A. Crystal structure of the mammalian Grb2 adaptor. *Science*. **1995**;268(5208):291-293. doi:10.1126/science.7716522
2. Kraskouskaya D, Duodu E, Arpin CC, Gunning PT. Progress towards the development of SH2 domain inhibitors. *Chem Soc Rev*. **2013**;42(8):3337-3370. doi:10.1039/c3cs35449k
3. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science*. **2003**;300(5618):445-452. doi:10.1126/science.1083653
4. Brown MT, Cooper JA. Regulation, substrates and functions of src. *Biochim Biophys Acta*. **1996**;1287(2-3):121-149. doi:10.1016/0304-419x(96)00003-0
5. Kawata T, Shevchenko A, Fukuzawa M, et al. SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. *Cell*. **1997**;89(6):909-916. doi:10.1016/s0092-8674(00)80276-7
6. Marengere LE, Pawson T. Structure and function of SH2 domains. *J Cell Sci Suppl*. **1994**;18:97-104. doi:10.1242/jcs.1994.supplement-18.14
7. Pawson T, Gish GD, Nash P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol*. **2001**;11(12):504-511. doi:10.1016/s0962-8924(01)02154-7
8. Nash PD. Why modules matter. *FEBS Lett*. **2012**;586(17):2572-2574. doi:10.1016/j.febslet.2012.04.049
9. Jaber Chehayeb R, Boggon TJ. SH2 Domain Binding: Diverse FLVRs of Partnership. *Front Endocrinol (Lausanne)*. **2020**;11:575220. Published 2020 Sep 18. doi:10.3389/fendo.2020.575220
10. Pawson T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell*. **2004**;116(2):191-203. doi:10.1016/s0092-8674(03)01077-8
11. Waksman G, Kominos D, Robertson SC, et al. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature*. **1992**;358(6388):646-653. doi:10.1038/358646a0
12. Machida K, Mayer BJ. The SH2 domain: versatile signaling module and pharmaceutical target. *Biochim Biophys Acta*. **2005**;1747(1):1-25. doi:10.1016/j.bbapap.2004.10.005
13. Sanches K, Caruso ÍP, Almeida FCL, Melo FA. NMR assignment of free 1H, 15N and 13C-Grb2-SH2 domain. *Biomol NMR Assign*. **2019**;13(2):295-298. doi:10.1007/s12104-019-09894-x
14. Kurochkina N, Guha U. SH3 domains: modules of protein-protein interactions. *Biophys Rev*. **2013**;5(1):29-39. doi:10.1007/s12551-012-0081-z
15. Gmeiner WH, Horita DA. Implications of SH3 domain structure and dynamics for protein regulation and drug design. *Cell Biochem Biophys*. **2001**;35(2):127-140. doi:10.1385/CBB:35:2:127
16. Kaneko T, Li L, Li SS. The SH3 domain–a family of versatile peptide- and protein-recognition module. *Front Biosci*. **2008**;13:4938-4952. Published 2008 May 1. doi:10.2741/3053
17. Mayer BJ. The discovery of modular binding domains: building blocks of cell signalling. *Nat Rev Mol Cell Biol*. **2015**;16(11):691-698. doi:10.1038/nrm4068
18. Musacchio A, Wilmanns M, Saraste M. Structure and function of the SH3 domain. *Prog Biophys Mol Biol*. **1994**;61(3):283-297. doi:10.1016/0079-6107(94)90003-5
19. Dalgarno DC, Botfield MC, Rickles RJ. SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers*. **1997**;43(5):383-400. doi:10.1002/(SICI)1097-0282(1997)43:5<383::AID-BIP4>3.0.CO;2-R
20. Ren R, Mayer BJ, Cicchetti P, Baltimore D. Identification of a ten-amino acid proline-rich SH3 binding site. *Science*. **1993**;259(5098):1157-1161. doi:10.1126/science.8438166
21. Yu H, Chen JK, Feng S, Dalgarno DC, Brauer AW, Schreiber SL. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell*. **1994**;76(5):933-945. doi:10.1016/0092-8674(94)90367-0
22. Feng S, Chen JK, Yu H, Simon JA, Schreiber SL. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*. **1994**;266(5188):1241-1247. doi:10.1126/science.7526465
23. Lim WA, Richards FM, Fox RO. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains [published correction appears in Nature 1995 Mar 2;374(6517):94]. *Nature*. **1994**;372(6504):375-379. doi:10.1038/372375a0

24. McDonald CB, Seldeen KL, Deegan BJ, Lewis MS, Farooq A. Grb2 adaptor undergoes conformational change upon dimerization. *Arch Biochem Biophys*. **2008**;475(1):25-35. doi:10.1016/j.abb.2008.04.008
25. Ahmed Z, Timsah Z, Suen KM, et al. Grb2 monomer-dimer equilibrium determines normal versus oncogenic function [published correction appears in Nat Commun. 2015;6:8007]. *Nat Commun*. **2015**;6:7354. Published 2015 Jun 24. doi:10.1038/ncomms8354
26. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. *Bioinformatics*. **1999**;15(4):327-332. doi:10.1093/bioinformatics/15.4.327
27. Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Pept Protein Res*. **1975**;7(6):445-459.
28. Clementi C, Nymeyer H, Onuchic JN. 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. *J Mol Biol*. **2000**;298(5):937-953. doi:10.1006/jmbi.2000.3693
29. Oliveira LC, Schug A, Onuchic JN. Geometrical features of the protein folding mechanism are a robust property of the energy landscape: a detailed investigation of several reduced models. *J Phys Chem B*. **2008**;112(19):6131-6136. doi:10.1021/jp0769835
30. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins*. **2009**;75(2):430-441. doi:10.1002/prot.22253
31. Noel JK, Levi M, Raghunathan M, et al. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput Biol*. **2016**;12(3):e1004794. Published 2016 Mar 10. doi:10.1371/journal.pcbi.1004794
32. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput*. **2008**;4(3):435-447. doi:10.1021/ct700301q
33. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. **1995**;91(1-3):43-56. doi:10.1016/0010-4655(95)00042-e
34. Swendsen RH. Modern methods of analyzing Monte Carlo computer simulations. *Physica A*. **1993**;194(1-4):53-62. doi:10.1016/0378-4371(93)90339-6
35. Oliveira AB Jr, Yang H, Whitford PC, Leite VBP. Distinguishing Biomolecular Pathways and Metastable States. *J Chem Theory Comput*. **2019**;15(11):6482-6490. doi:10.1021/acs.jctc.9b00704
36. Tejada E, Minghim R, Nonato LG. On improved projection techniques to support visual exploration of multi-dimensional data sets. *Inf Vis*. **2003**;2(4):218-231. doi:10.1057/palgrave.ivs.9500054
37. Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. **1999**;112:531-552. doi:10.1385/1-59259-584-7:531
38. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem*. **2000**;287(2):252-260. doi:10.1006/abio.2000.4880
39. Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins*. **1999**;35(3):307-312.
40. Lakowicz JR. Instrumentation for Fluorescence Spectroscopy. In: Principles of Fluorescence Spectroscopy. *Springer US*; **1999**:25-61.
41. Pérez HA, Disalvo A, Frías MLÁ. Effect of cholesterol on the surface polarity and hydration of lipid interphases as measured by Laurdan fluorescence: New insights. *Colloids Surf B Biointerfaces*. **2019**;178:346-351. doi:10.1016/j.colsurfb.2019.03.022
42. Pulavarti SVSRK, Maguire JB, Yuen S, et al. From Protein Design to the Energy Landscape of a Cold Unfolding Protein. *J Phys Chem B*. **2022**;126(6):1212-1231. doi:10.1021/acs.jpcb.1c10750
43. Ochi S, Iiyama M, Oda M. Interdomain interactions in Grb2 revealed by the conformational stability and CD28 binding analysis. *Biophys Chem*. **2022**;284:106792. doi:10.1016/j.bpc.2022.106792
44. Kato A, Takagi T. Formation of intermolecular .beta.-sheet structure during heat denaturation of ovalbumin. *J Agric Food Chem*. **1988**;36(6):1156-1159. doi:10.1021/jf00084a007
45. Lefèvre T, Subirade M. Formation of intermolecular beta-sheet structures: a phenomenon relevant to protein film structure at oil-water interfaces of emulsions. *J Colloid Interface Sci*. **2003**;263(1):59-67. doi:10.1016/s0021-9797(03)00252-2
46. Dobson CM. Protein folding and misfolding. *Nature*. **2003**;426(6968):884-890. doi:10.1038/nature02261
47. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol*. **2007**;8(4):319-330. doi:10.1038/nrm2144
48. Lin CC, Melo FA, Ghosh R, et al. Inhibition of basal FGF receptor signaling by dimeric Grb2. *Cell*. **2012**;149(7):1514-1524. doi:10.1016/j.cell.2012.04.033
49. Ahmed Z, Lin CC, Suen KM, et al. Grb2 controls phosphorylation of FGFR2 by inhibiting receptor kinase and Shp2 phosphatase activity. *J Cell Biol*. **2013**;200(4):493-504. doi:10.1083/jcb.201204106
50. Sanches K, Dias RVR, da Silva PH, et al. Grb2 dimer interacts with Coumarin through SH2 domains: A combined experimental and molecular modeling study. *Heliyon*. **2019**;5(11):e02869. Published 2019 Nov 27. doi:10.1016/j.heliyon.2019.e02869
51. Sanches K, Dias RVR, da Silva PH, et al. Thermodynamic profile and molecular modeling of the interaction between Grb2 dimer and flavonoids Rutin and Morin. *J Mol Struct*. **2021**;1234(130164):130164. doi:10.1016/j.molstruc.2021.130164