

RESSALVA

Atendendo solicitação do autor,
o texto completo desta tese será
disponibilizado somente a partir
de 27/01/2025



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 últimos 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 científico que vivemos. E Muliro, apresentar o zeldinha nos últimos meses de doutorado foi sacanagem !

Gi e Naza, muito obrigado por tudo! Cada discussão sempre tem uma tonelada de idéias, é até difícil 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 solícitos, 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 transição 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 resíduo 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	<p>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.. . . .</p>	28
6	<p>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 (intermediário/Nativo) é a mais custosa energeticamente em relação as outras transições</p>	31
7	<p>Superfície de energia e clusterização de estruturas. A) Superfície 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.</p>	32

8	<p>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.</p>	34
9	<p>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 (P_{ij}). passando de poucos contatos formados e domínios desdobrados (azul) para contatos formados e domínios dobrados (vermelho). . . .</p>	35
10	<p>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 (T_s). 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.</p>	36
11	<p>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-β 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.</p>	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 pelo 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 $C_{TERMINAL}$ 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	<p>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.</p>	56
22	<p>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.</p>	58
23	<p>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.</p>	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

C α Carbono Alfa

GRB2 *Growth Factor Receptor Bound protein 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 ϕ	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	<i>Growth Factor Receptor Bound 2 (GRB2)</i>	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	Glicoproteínas	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

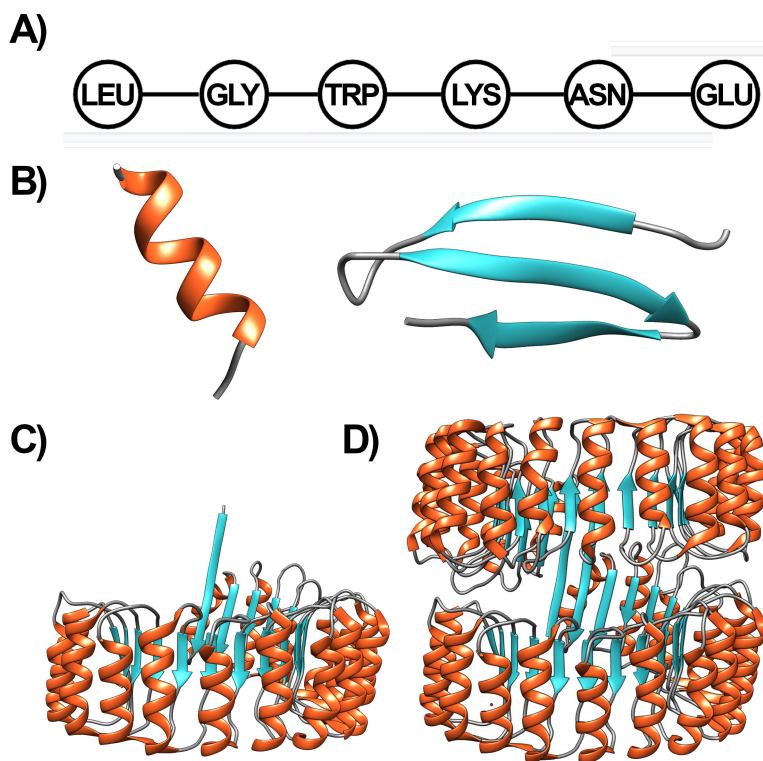
1.1 Estrutura e função das proteínas

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 ($C\alpha$) fazendo quatro ligações covalentes: uma com um átomo de hidrogênio (H), outra com um grupo amina (NH_2), 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 conseqüentemente 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:

$$F(Q) = E(Q) - TS(Q) \quad (1)$$

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:

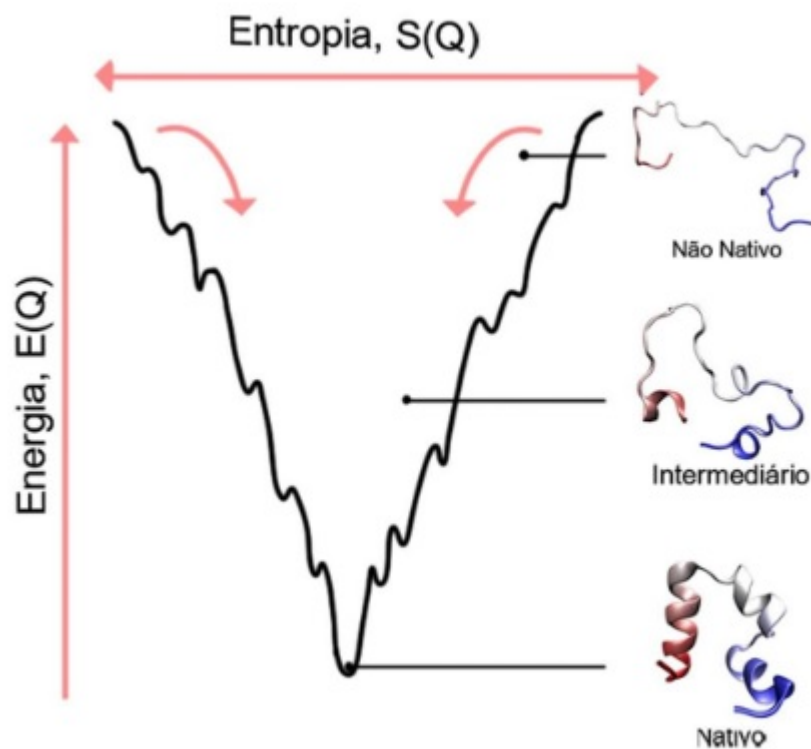
I) a proteína GRB2, uma proteína adaptadora com multidomínios, onde são avaliados os mecanismos que se relacionam ao seu enovelamento. Essa é uma proteína de interesse devido à sua organização estrutural em três diferentes domínios que se contrastam aos processos biológicos em que ela aparece envolvida, 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).

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 influencia aumentando a importância de alguns resíduos para o enovelamento.

Os resultados encontrados indicam que com uma modelagem simples, como é o $C\alpha$ 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 ANFENSEN, 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 physicochemistry*, 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.; WILMANN, 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.