

Trabalho de Conclusão de Curso

Curso de Graduação em Física

**ESTUDO DO RELEVO DE ENERGIA POTENCIAL DA PROTEÍNA COM MUDANÇA
DE ENOVELAMENTO KAIB**

Giovana Trevejo

Prof. Dr. Florian Steffen Günther

Rio Claro (SP)

2025

UNIVERSIDADE ESTADUAL PAULISTA
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

GIOVANA TREVEJO

ESTUDO DO RELEVO DE ENERGIA POTENCIAL DA
PROTEÍNA COM MUDANÇA DE ENOVELAMENTO KAIB

Trabalho de Conclusão de Curso apresentado
ao Instituto de Geociências e Ciências Exatas -
Câmpus de Rio Claro, da Universidade
Estadual Paulista Júlio de Mesquita Filho, para
obtenção do grau de Bacharel em Física.

Rio Claro - SP
2025

T812e	Trevejo, Giovana Estudo do relevo de energia potencial da proteína com mudança de enovelamento KaiB / Giovana Trevejo. -- Rio Claro, 2025 51 p. Trabalho de conclusão de curso (Bacharelado - Física) - Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro Orientador: Florian Steffen Günther Coorientador: Vitor Barbanti Pereira Leite 1. Enovelamento de proteínas. 2. Proteína KaiB. 3. Modelos baseados em estrutura. 4. Dinâmica molecular. 5. ELViM. I. Título.
-------	---

GIOVANA TREVEJO

ESTUDO DO RELEVO DE ENERGIA POTENCIAL DA
PROTEÍNA COM MUDANÇA DE ENOVELAMENTO KAIB

Trabalho de Conclusão de Curso apresentado
ao Instituto de Geociências e Ciências Exatas -
Câmpus de Rio Claro, da Universidade
Estadual Paulista Júlio de Mesquita Filho, para
obtenção do grau de Bacharel em Física.

Comissão Examinadora

Prof. Dr. Florian Steffen Günther (orientador)

Prof. Dr. Ricardo Paupitz Barbosa dos Santos

Prof. Dr. Rafael Jesus Gonçalves Rubira

Rio Claro, 30 de Outubro de 2025.

Assinatura do(a) aluno(a)

Assinatura do(a) orientador(a)

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo nº 128740/2024-8), pelo apoio financeiro que possibilitou a realização desta pesquisa.

Ao meu orientador, professor Florian, pela orientação, paciência e pelos aprendizados compartilhados ao longo deste trabalho.

Ao professor Vitor e ao grupo de pesquisa, pelas discussões que contribuíram para o andamento do trabalho.

Aos membros da banca por aceitarem o convite e dedicarem seu tempo à avaliação deste trabalho.

À minha família, em especial à minha mãe Daniela, meu pai José Antonio e minha irmã Lívia, pelo apoio, paciência e incentivo, e por sempre me ouvir, incluindo todos os desabafos, reclamações e histórias de simulações que deram errado, muitas vezes mais de uma vez no mesmo dia.

Aos amigos que estiveram presentes durante a graduação, pela companhia, pelas conversas e pela leveza que tornaram esse caminho mais fácil de percorrer.

RESUMO

O processo de enovelamento de proteínas, responsável por levá-las à sua conformação funcional, é fundamental para a compreensão de diversos mecanismos biológicos. Embora muitas proteínas apresentem uma única estrutura nativa, algumas exibem múltiplos estados conformacionais funcionalmente relevantes, como ocorre com a proteína KaiB, envolvida na regulação do ciclo circadiano de cianobactérias. Este trabalho teve como objetivo estudar o relevo de energia potencial associado às transições conformacionais da KaiB, utilizando modelos computacionais baseados em estrutura. Inicialmente, empregou-se o domínio SH3 como sistema modelo, por apresentar um enovelamento bem caracterizado e cooperativo, permitindo validar a metodologia. Foram realizadas simulações em diferentes temperaturas para analisar as coordenadas de reação $Q_{contato}$, q_w e RMSD, possibilitando identificar a temperatura de transição e caracterizar as flutuações estruturais, tanto da SH3 quanto dos estados funcionais da KaiB, gs e fs. A análise das coordenadas de reação, dos desvios padrão e do calor específico permitiu estimar as respectivas temperaturas de enovelamento, de aproximadamente 0.88 para a gsKaiB e 0.78 para a fsKaiB. Posteriormente, aplicou-se o método ELViM para projetar o espaço conformacional, permitindo visualizar conformações intermediárias e interpretar os caminhos estruturais durante a transição. Os resultados indicam que a gsKaiB apresenta maior estabilidade térmica em comparação à fsKaiB, o que é coerente com seus distintos papéis funcionais no ciclo circadiano. Este estudo fornece uma base para a aplicação futura do modelo *dual basin*, que permitirá simular transições entre os dois estados conformacionais da KaiB de forma mais realista.

Palavras-chave: enovelamento de proteínas; KaiB; modelos baseados em estrutura; ELViM.

ABSTRACT

The protein folding process, responsible for leading proteins to their functional conformation, is fundamental to the understanding of various biological mechanisms. Although many proteins have a single native structure, some exhibit multiple functionally relevant conformational states, as is the case with KaiB, a protein involved in regulating the circadian rhythm of cyanobacteria. This work aimed at studying the potential energy landscape associated with KaiB's conformational transitions using structure-based computational models. Initially, the SH3 domain was used as a model system because it displays well-characterized and cooperative folding behavior, allowing to validate the methodology. Simulations were performed at different temperatures to analyze the reaction coordinates $Q_{contact}$, q_w and RMSD, enabling us to identify the transition temperature and characterize the structural fluctuations of both SH3 and KaiB's functional states, gs and fs. Analysis of the reaction coordinates, standard deviations, and specific heat allowed estimation of the respective folding temperatures, approximately 0.88 for gsKaiB and 0.78 for fsKaiB. Subsequently, the ELViM method was applied to project the conformational space, allowing the visualization of intermediate conformations and the interpretation of structural pathways during the transition. The results indicate that gsKaiB exhibits greater thermal stability compared to fsKaiB, consistent with their distinct functional roles in the circadian rhythm. This study provides a foundation for future application of the dual basin model, which will allow for more realistic simulation of transitions between the two conformational states of KaiB.

Keywords: protein folding; KaiB; structure-based models; ELViM.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estrutura geral de um aminoácido, destacando o carbono α (esfera central), grupo amina (NH_2), carboxila (COOH) e cadeia lateral (R).....	10
Figura 2 - Níveis estruturais de uma proteína.....	10
Figura 3 - Representação esquemática de dois tipos comuns de estrutura secundária de proteínas: (A) hélice- α e (B) folha- β	11
Figura 4 - Representação da estrutura primária (A) e terciária (B) do domínio SH3 (PDB id: 1fmk).....	14
Figura 5 - Representação da estrutura primária (A), secundária (B) e terciária (C) das duas conformações da proteína KaiB: gsKaiB (esquerda) e fsKaiB (direita). As cores representam os elementos de estrutura secundária: α -hélices (rosa), folhas β (verde) e regiões desordenadas (cinza).....	15
Figura 6 - Representação esquemática dos diferentes termos que compõem a energia potencial em modelos coarse-grained.....	18
Figura 7 - Esquema ilustrativo do shadow contacts map.....	21
Figura 8 - Representação esquemática dos potenciais utilizados.....	23
Figura 9 - Exemplo de duas conformações (k e l) nas quais são mostradas as distâncias internas entre os resíduos 1, 7, 13, 33 e 60.....	25
Figura 10 - Análise da coordenada de reação Q_{contato} ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente.....	31
Figura 11 - Análise da coordenada de reação q_w ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente.....	31
Figura 12 - Análise da coordenada de reação RMSD ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente.....	31
Figura 13 - Gráfico das médias das coordenadas de reação em função da temperatura para a SH3.....	32
Figura 14 - Variação do calor específico da SH3 em função da temperatura.....	33
Figura 15 - Projeção bidimensional do ELViM da SH3, na temperatura 1.09, como função das coordenadas de reação Q_{contato} (A), q_w (B) e RMSD (C)..	34
Figura 16 - Gráfico das médias das coordenadas de reação em função da	

temperatura para a conformação gsKaiB.....	36
Figura 17 - Análise da coordenada de reação Qcontato ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente.....	37
Figura 18 - Análise da coordenada de reação qw ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente.....	37
Figura 19 - Análise da coordenada de reação RMSD ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente.....	38
Figura 20 - Variação do calor específico da conformação gsKaiB em função da temperatura.....	38
Figura 21 - Projeção bidimensional do ELViM da gsKaiB, na temperatura 0.88, como função das coordenadas de reação Qcontato (A), qw (B) e RMSD (C)..	39
Figura 22 - Gráfico das médias das coordenadas de reação em função da temperatura para a conformação fsKaiB.....	40
Figura 23 - Análise da coordenada de reação Qcontato ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente.....	41
Figura 24 - Análise da coordenada de reação qw ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente.....	41
Figura 25 - Análise da coordenada de reação RMSD ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente.....	42
Figura 26 - Variação do calor específico da proteína fsKaiB em função da temperatura.....	43
Figura 27 - Projeção bidimensional do ELViM da fsKaiB, na temperatura 0.78, como função das coordenadas de reação Qcontato (A), qw (B) e RMSD (C), respectivamente.....	44

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1 Estrutura e composição das proteínas.....	9
1.2 O problema do enovelamento de proteínas.....	11
1.3 A importância de modelos simplificados.....	13
1.4 Domínio SH3 como sistema modelo.....	14
1.5 Proteína KaiB: estrutura, função e dinâmica conformacional.....	14
1.6 Objetivos do trabalho.....	16
2. METODOLOGIA.....	17
2.1 Modelos baseados em estrutura.....	17
2.1.1 Potencial de Lennard-Jones modificado.....	19
2.1.2 Potencial Gaussiano.....	21
2.1.3 Potencial com duplo mínimo.....	22
2.2 Coordenadas de reação.....	23
2.2.1 Calor específico.....	26
2.3 Energy Landscape Visualization Method.....	27
2.4 Detalhes computacionais.....	28
2.5 Análise das trajetórias.....	29
3. RESULTADOS.....	30
3.1 Análise do domínio SH3.....	30
3.2 Análise da proteína KaiB.....	35
3.2.1 Conformação gsKaiB.....	36
3.2.2 Conformação fsKaiB.....	40
4. DISCUSSÃO.....	46
5. CONCLUSÃO.....	48
REFERÊNCIAS.....	49

1. INTRODUÇÃO

1.1 Estrutura e composição das proteínas

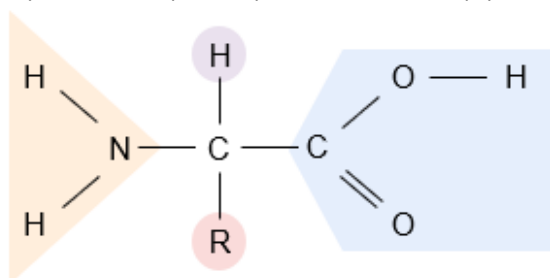
As proteínas são heteropolímeros lineares formados por resíduos de aminoácidos (Alberts et al., 2017), ou seja, aminoácidos que já foram incorporados à cadeia polipeptídica por ligações peptídicas. Existem 20 tipos distintos de aminoácidos comumente encontrados na natureza (tabela 1), cada um composto por um carbono central (carbono α) ligado a quatro grupos: um átomo de hidrogênio, um grupo amina (NH_2), um grupo carboxila (COOH) e uma cadeia lateral (grupo R) (Figura 1) (Alberts et al., 2017). As cadeias laterais, que não participam das ligações peptídicas, atribuem propriedades únicas a cada aminoácido, variando em estrutura, tamanho, carga elétrica e hidrofobicidade (Alberts et al., 2017).

Tabela 1 - Nomes e abreviações dos 20 aminoácidos comumente encontrados na natureza

Aminoácido	Abreviação	
Ácido aspártico	Asp	D
Ácido glutâmico	Glu	E
Arginina	Arg	R
Lisina	Lys	K
Histidina	His	H
Asparagina	Asn	N
Glutamina	Gln	Q
Serina	Ser	S
Treonina	Thr	T
Tirosina	Tyr	Y
Alanina	Ala	A
Glicina	Gly	G
Valina	Val	V
Leucina	Leu	L
Isoleucina	Ile	I
Prolina	Pro	P
Fenilalanina	Phe	F
Metionina	Met	M
Triptofano	Trp	W
Cisteína	Cys	C

Fonte: Adaptado de ALBERTS et al. (2017, p. 109).

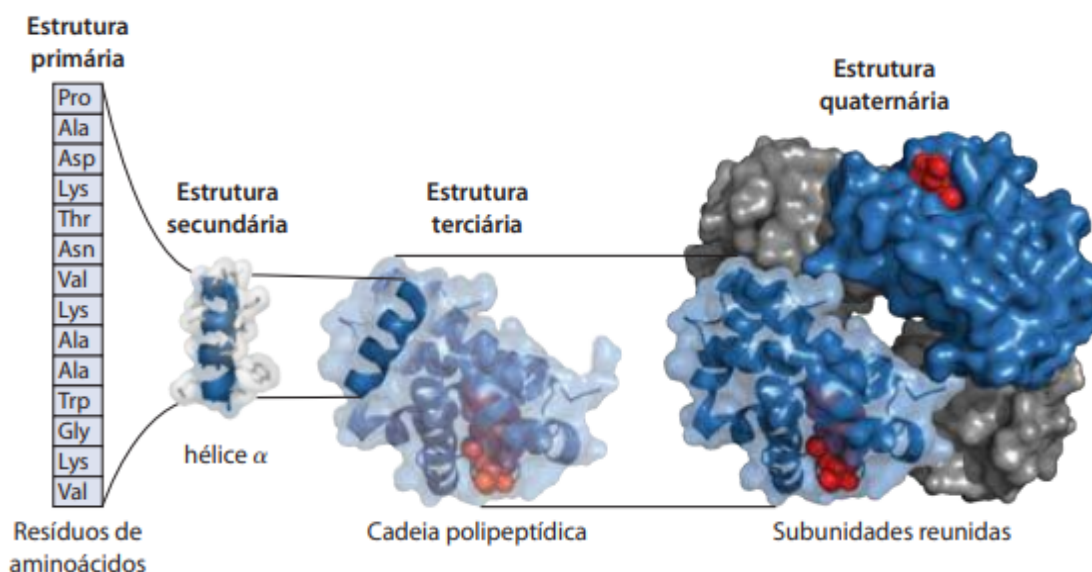
Figura 1 - Estrutura geral de um aminoácido, destacando o carbono α (esfera central), grupo amina (NH_2), carboxila (COOH) e cadeia lateral (R)



Fonte: Elaboração própria.

A estrutura das proteínas é organizada de forma hierárquica, composta por quatro níveis interdependentes (Figura 2). A estrutura primária corresponde à sequência linear de aminoácidos determinada geneticamente (Nelson; Cox, 2014).

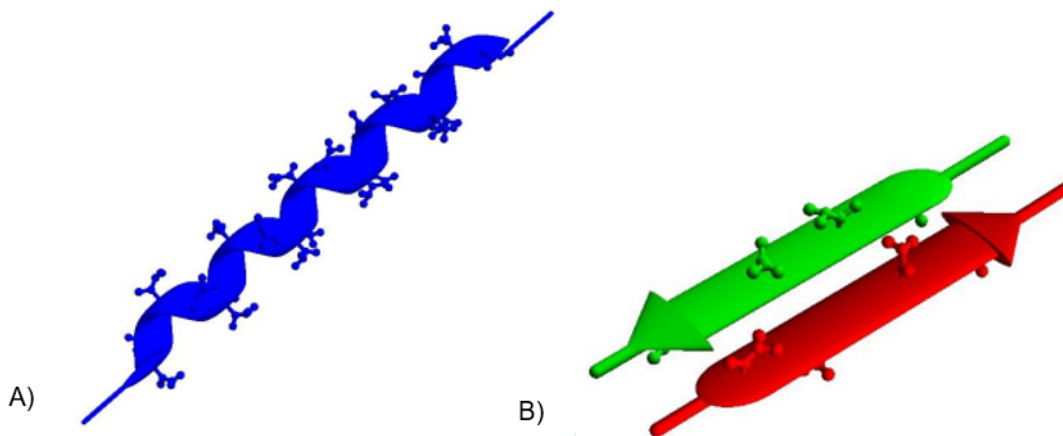
Figura 2 - Níveis estruturais de uma proteína



Fonte: Retirado de Nelson (2014, p. 96)

A estrutura secundária refere-se ao arranjo espacial dos átomos da cadeia principal, sem considerar a posição das cadeias laterais (Nelson; Cox, 2014). Alguns tipos de estruturas secundárias são especialmente estáveis e bastante presentes em proteínas, sendo as hélices α e as conformações β as mais conhecidas (Figura 3) (Nelson; Cox, 2014).

Figura 3 - Representação esquemática de dois tipos comuns de estrutura secundária de proteínas: (A) hélice- α e (B) folha- β .



Fonte: Adaptado de Günther (2013, p. 12)

A estrutura terciária corresponde ao arranjo tridimensional completo de todos os átomos de uma proteína, envolvendo interações entre resíduos que podem estar distantes na sequência. Essas interações determinam a conformação final da proteína enovelada (Nelson; Cox, 2014). Quando duas ou mais cadeias polipeptídicas, cada uma com sua própria estrutura terciária, se associam funcionalmente, forma-se a estrutura quaternária (Nelson; Cox, 2014).

O entendimento dessas diferentes camadas estruturais é essencial para compreender o processo de enovelamento e, conseqüentemente, a função biológica das proteínas (Dobson, 2003).

1.2 O problema do enovelamento de proteínas

O enovelamento proteico é o processo pelo qual uma cadeia linear de aminoácidos adquire sua conformação tridimensional funcional, conhecida como estado nativo (Nelson; Cox, 2014 e Dobson, 2003). Um dos grandes desafios da biofísica é compreender como a sequência de aminoácidos guia esse processo de forma eficiente e reprodutível (Dobson, 2003).

Na década de 60, Christian Anfinsen demonstrou que a sequência de aminoácidos de uma proteína contém a informação necessária para definir sua conformação tridimensional nativa, correspondente ao estado de menor energia livre de Gibbs (Anfinsen, 1973). Em experimentos com a ribonuclease, verificou-se que, após a desnaturação, a molécula recuperava espontaneamente sua forma funcional,

indicando que o enovelamento é determinado pela própria sequência. Esses resultados levaram à formulação da “Hipótese Termodinâmica”, segundo a qual o enovelamento proteico corresponde a um processo de minimização da energia livre, regido pela estrutura primária, mas intrinsecamente complexo devido ao grande número de conformações possíveis (Anfinsen, 1973).

Apesar dessa formulação indicar o estado final do processo, ela não esclarece de que forma a proteína alcança sua conformação nativa em escalas de tempo compatíveis com os sistemas biológicos. Essa limitação levantou a necessidade de compreender os caminhos de enovelamento, e é nesse contexto que surge o Paradoxo de Levinthal, proposto por Cyrus Levinthal, em 1968. Ele argumentou que, se uma proteína testasse aleatoriamente todas as possíveis conformações até encontrar a estrutura nativa, o tempo necessário seria extremamente longo, mesmo para cadeias pequenas.

Como exemplificado no estudo de Contessoto et al., uma proteína com 100 resíduos teria cerca de 2^{100} conformações possíveis, aproximadamente 10^{30} conformações, uma vez que cada resíduo pode adotar múltiplas orientações no espaço; se cada conformação fosse testada em 1 picosegundo, o tempo total necessário para acessar todas as 10^{30} conformações seria de 10^{18} segundos, o que equivale 10^{10} anos, comparável à idade do universo. No entanto, sabe-se que proteínas pequenas podem se enovelar em questão de segundos. Isso indica que o enovelamento não ocorre ao acaso, mas segue trajetórias favorecidas pela superfície de energia, guiado por propriedades termodinâmicas e cinética da cadeia polipeptídica (Contessoto et al., 2018).

A superfície de energia associada ao processo é multidimensional, com inúmeros graus de liberdade proporcional ao número de átomos da molécula. Isso dificulta tanto a representação visual quanto a amostragem eficiente dos caminhos que a proteína pode seguir até atingir seu estado nativo (Oliveira Junior., 2017).

Esses desafios continuam sendo um tema central na ciência contemporânea. A relevância desse campo foi recentemente reconhecida pelo Prêmio Nobel de Química de 2024, concedido a David Baker, pelo desenvolvimento de métodos para projetar proteínas inteiramente novas, e a Demis Hassabis e John Jumper, pelo *AlphaFold2*, um modelo de inteligência artificial capaz de prever estruturas tridimensionais de proteínas com alta precisão (The Nobel Prize, 2024). Esse

reconhecimento evidencia tanto a capacidade de projetar proteínas inéditas quanto o avanço em resolver o problema de prever o enovelamento a partir da sequência de aminoácidos, destacando a importância de compreender as estruturas e funções proteicas como um dos grandes desafios em aberto da biologia molecular moderna.

1.3 A importância de modelos simplificados

A escolha do modelo a ser adotado está relacionada à questão específica a ser investigada. É importante considerar perguntas como: Quais simplificações podem ser feitas sem comprometer a validade das conclusões? Quais são as limitações computacionais envolvidas? Quais aspectos do fenômeno são essenciais para a análise? Essas perguntas possibilitam medir a complexidade e aplicabilidade do modelo escolhido.

Incluir muitos detalhes no modelo tende a aumentar significativamente o custo computacional, o que nem sempre é viável para certos tipos de análise. Em contrapartida, um modelo excessivamente simplificado pode não ser capaz de explicar adequadamente os aspectos observáveis experimentalmente (Contessoto et al., 2018). Assim, a decisão deve considerar cuidadosamente os objetivos do estudo e os aspectos mais relevantes do sistema. Um bom modelo é aquele que consegue representar com fidelidade os pontos centrais do problema, sem deixar de ser viável computacionalmente.

Simulações computacionais baseadas em modelos simplificados possibilitam a análise de uma grande variedade de parâmetros dos sistemas com um custo computacional mais baixo. Esse tipo de abordagem tem se mostrado eficaz para investigar e compreender o processo de enovelamento de proteínas (Oliveira Junior, 2017).

Nesse contexto, a análise do enovelamento de proteínas adota uma abordagem teórica que emprega modelos simplificados para entender os princípios fundamentais que determinam esse processo. Os modelos baseados em estrutura e os potenciais associados a eles, que serão abordados em mais detalhes na metodologia, são uma estratégia simplificada que usa a configuração nativa da proteína como referência para definir as interações entre seus componentes, partindo do princípio de que a estrutura nativa contém todas as informações necessárias para determinar essas interações (Oliveira Junior, 2017). Esse tipo de modelagem é eficaz para explorar aspectos fundamentais do enovelamento, como

as superfícies de energia associadas ao processo, os caminhos preferenciais que as proteínas seguem para atingir o estado nativo e os efeitos de mutações pontuais sobre o processo (Oliveira Junior, 2017).

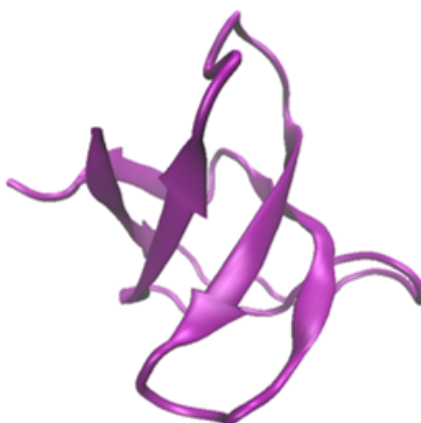
1.4 Domínio SH3 como sistema modelo

Como parte do processo de familiarização com simulações de dinâmica molecular, foi utilizado o domínio SH3, PDB id: 1fmk (Figura 4). Esse domínio foi escolhido devido à sua estrutura compacta e bem caracterizada, o que facilita a validação de métodos de simulação (Hubner; Edmonds; Shakhnovich, 2005).

Figura 4 - Representação da estrutura primária (A) e terciária (B) do domínio SH3 (PDB id: 1fmk)

A) **TTFVALYDYESRTETDLSFKKGERLQIVNNTGDDWWLAHSLSTGQTGYIPSNYVAPSD**

B)



Fonte: Elaboração própria

O domínio SH3 serviu como referência para estudar enovelamento e estabilidade estrutural, permitindo aprendizado prático das ferramentas computacionais, execução de dinâmicas moleculares e análise da evolução estrutural ao longo do tempo, consolidando o domínio das metodologias utilizadas posteriormente na proteína KaiB.

1.5 Proteína KaiB: estrutura, função e dinâmica conformacional

O ciclo circadiano é um mecanismo biológico fundamental que regula os ritmos internos da maioria dos seres vivos e ajusta o metabolismo celular conforme a alternância entre luz e escuridão (Cohen; Golden, 2015).

Em cianobactérias, esse ritmo é controlado por um sistema molecular

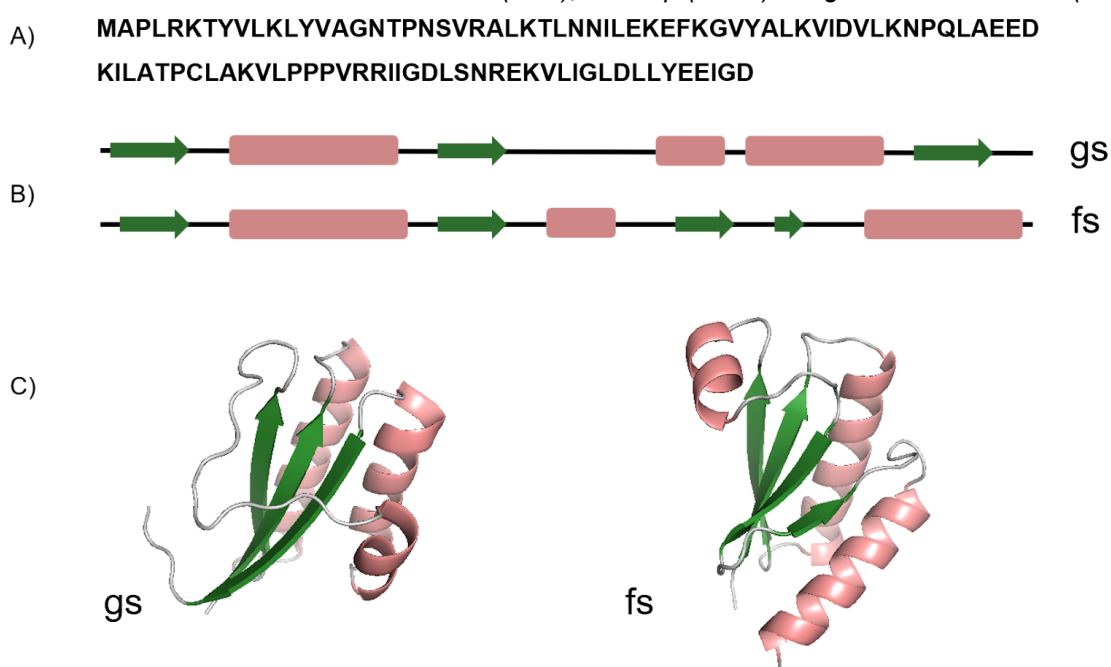
relativamente simples composto pelas proteínas KaiA, KaiB e KaiC (Cohen; Golden, 2015). Esse conjunto regula a fosforilação cíclica da proteína KaiC, que atua como marcador interno de tempo (Chang et al., 2015).

Durante o dia subjetivo, a KaiA estimula a autofosforilação de KaiC. Por outro lado, a KaiB exerce uma função reguladora inibitória, sendo responsável por iniciar a fase oposta do ciclo. Ela interage com a KaiC fosforilada e bloqueia a ação da KaiA, promovendo a desfosforilação da KaiC durante o período noturno (Chang et al., 2015).

Um dos aspectos interessantes da KaiB é a sua capacidade de alternar entre dois estados conformacionais distintos: gsKaiB, PDB id: 1vgl, e fsKaiB, PDB id: 5jyt (Figura 5). Essa transição envolve a quebra e a formação de diversas interações intramoleculares. No modelo *coarse-grained*, a gsKaiB apresenta 242 contatos nativos, enquanto a fsKaiB apresenta 248. Esse resultado sugere que a transição não altera drasticamente a quantidade de interações, mas modifica quais contatos estão presentes e, conseqüentemente, a estabilidade relativa de cada conformação.

Durante o ciclo circadiano, a forma estrutural predominante é a gsKaiB, que gradualmente transita para a fsKaiB. Essa mudança introduz um atraso temporal essencial para a sincronização do ciclo com a rotação da Terra (Cohen; Golden, 2015).

Figura 5 - Representação da estrutura primária (A), secundária (B) e terciária (C) das duas conformações da proteína KaiB: gsKaiB (esquerda) e fsKaiB (direita). As cores representam os elementos de estrutura secundária: α -hélices (rosa), folhas β (verde) e regiões desordenadas (cinza)



Fonte: Elaboração própria

Devido à sua capacidade de alternar entre conformações estáveis, a KaiB é um bom modelo para estudos computacionais de mudanças estruturais em proteínas. Compreender essa dinâmica é essencial para elucidar os mecanismos moleculares dos ritmos biológicos e pode ter aplicações futuras em biotecnologia e otimização de processos fotossintéticos. Para investigar suas interações com outras proteínas do sistema, modelos teóricos simplificados têm sido aplicados, permitindo identificar fatores estruturais que influenciam sua função (Cohen; Golden, 2015).

1.6 Objetivos do trabalho

- 1) Investigar o processo de enovelamento de proteínas utilizando o domínio SH3 como sistema modelo, de modo a compreender seus mecanismos e, ao mesmo tempo, familiarizar-se com os procedimentos e ferramentas computacionais aplicados no estudo de transições conformacionais.
- 2) Realizar uma dinâmica molecular das conformações gsKaiB e fsKaiB em diferentes temperaturas para identificar suas respectivas temperaturas de transição e analisar as diferenças de estabilidade entre os estados.
- 3) Construir e aplicar um modelo *dual basin* para a KaiB, permitindo estudar simultaneamente as duas conformações e investigar a transição da gsKaiB para a fsKaiB.

2. METODOLOGIA

2.1 Modelos baseados em estrutura

O Modelo Baseado em Estrutura define a energia do sistema a partir das propriedades geométricas da conformação nativa fornecida, partindo da premissa de que essa conformação contém todas as informações mínimas necessárias para descrever o comportamento termodinâmico e cinético do processo de enovelamento de proteína (Clementi; Nymeyer; Onuchic, 2000).

Nessa modelagem, tanto o potencial de interação quanto a representação estrutural são simplificados. Em modelos atomísticos, cada átomo da proteína é representado explicitamente, o que aumenta a precisão, mas demanda elevado custo computacional. Já nos modelos *coarse-grained*, grupos de átomos são reunidos em partículas efetivas, reduzindo o número de graus de liberdade e permitindo explorar escalas de tempo e tamanho inacessíveis a modelos atomísticos (Contessoto et al., 2018).

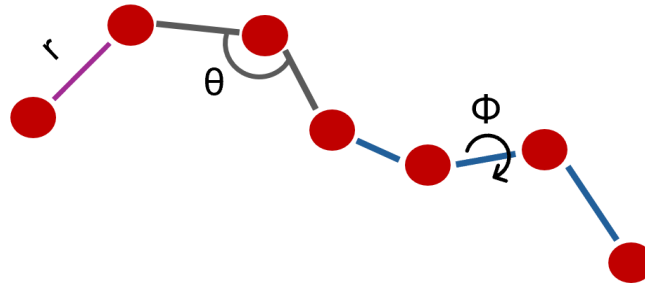
Um dos níveis mais comuns de *coarse-graining* em proteínas é o modelo de carbono alfa (C α), no qual cada resíduo de aminoácido é representado por uma esfera centrada no átomo C α (Oliveira Junior, 2017). Esse modelo foca exclusivamente nas posições desses átomos, eliminando detalhes atômicos dos grupos laterais e outros componentes da cadeia, o que contribui para uma descrição mais enxuta da estrutura (Clementi; Nymeyer; Onuchic, 2000).

As contribuições à energia total do sistema incluem interações covalentes e interações não covalentes (Oliveira; Schug; Onuchic, 2008). A energia potencial total do sistema é, portanto, dada por

$$E_{total} = E_{covalentes} + E_{angulares} + E_{diedrais} + E_{contatos} + E_{não-contatos} \quad (1)$$

Esses diferentes contribuintes estão representados esquematicamente na Figura 6, que ilustra como cada termo está associado a uma característica estrutural da proteína.

Figura 6 - Representação esquemática dos diferentes termos que compõem a energia potencial em modelos coarse-grained



Fonte: Elaboração própria

O termo *covalentes* considera as distâncias entre dois $C\alpha$ ligados covalentemente, sendo representado da forma

$$E_{covalentes} = \sum_{covalentes} \frac{1}{2} \epsilon_r (r - r_0)^2 \quad (2)$$

em que r é a distância entre os átomos, r_0 é a distância de referência na estrutura nativa e ϵ_r é a constante de ligação (Oliveira; Schug; Onuchic, 2008).

A energia angular refere-se ao ângulo formado por três $C\alpha$ em sequência e é escrita da forma

$$E_{angulares} = \sum_{angulares} \frac{1}{2} \epsilon_\theta (\theta - \theta_0)^2 \quad (3)$$

onde θ é o ângulo entre duas ligações adjacentes, θ_0 é o ângulo de referência na estrutura nativa e ϵ_θ é a constante angular (Oliveira; Schug; Onuchic, 2008).

O terceiro termo, *diedrais*, é o ângulo formado entre dois planos definidos por quatro átomos conectados em sequência. É dado pela expressão

$$E_{diedrais} = \sum_{diedrais} \left\{ \epsilon_\phi [1 - \cos(\phi - \phi_0)] + \frac{1}{2} \epsilon_\phi [1 - \cos(3(\phi - \phi_0))] \right\} \quad (4)$$

no qual, ϕ é o ângulo entre os dois planos, ϕ_0 é o ângulo dos planos no estado nativo e ϵ_ϕ é a constante do diedro (Oliveira; Schug; Onuchic, 2008).

As interações não covalentes incluem dois termos distintos: os contatos nativos e os não-contatos, ou contatos excluídos. O termo *não-contatos* (equação 5) corresponde a pares de resíduos que não estão em contato na estrutura nativa. Este termo é puramente repulsivo e impede que resíduos não nativos colapsem ou se sobreponham, preservando o volume excluído da proteína (Retamal-Farfán et al., 2023).

$$\sum_{\text{não-contatos}} \epsilon'_{LJ} \left(\frac{4\text{\AA}}{r_{ij}} \right)^{12} \quad (5)$$

As constantes energéticas ϵ_r , ϵ_θ , ϵ_ϕ e ϵ'_{LJ} foram definidas com valores de 200, 40, 1 e 1, respectivamente, em unidades de ϵ_{LJ} . Este último corresponde à energia de interação entre um par de contatos nativos, com valor aproximado de 1 kJ/mol. Esses parâmetros são fixos e foram estabelecidos no modelo proposto por Clementi et al. (2000).

Os contatos nativos correspondem a pares de resíduos próximos na estrutura nativa e são representados por um potencial de Lennard-Jones modificado, que inclui termos atrativos e repulsivos, garantindo a estabilidade da distância observada na conformação de referência. Essa questão será melhor explicada na seção a seguir.

2.1.1 Potencial de Lennard-Jones modificado

O potencial de Lennard-Jones é uma forma matemática usada para descrever interações não covalentes entre átomos. Ele combina uma parte repulsiva, que impede que os átomos fiquem muito próximos, e uma parte atrativa, que mantém os átomos a uma certa distância estável.

Na sua forma tradicional, a atração decai com r^{-6} , mas em modelos simplificados como o que utilizamos aqui, é mais comum empregar uma versão modificada, na qual a atração decai com r^{-10} . A principal diferença é que o potencial modificado é mais suave e evita valores extremos de energia em distâncias curtas. Isso o torna mais estável e viável para simulações de dinâmica molecular, especialmente em modelos *coarse-grained*, onde cada partícula representa um grupo de átomos. A equação usada neste trabalho é:

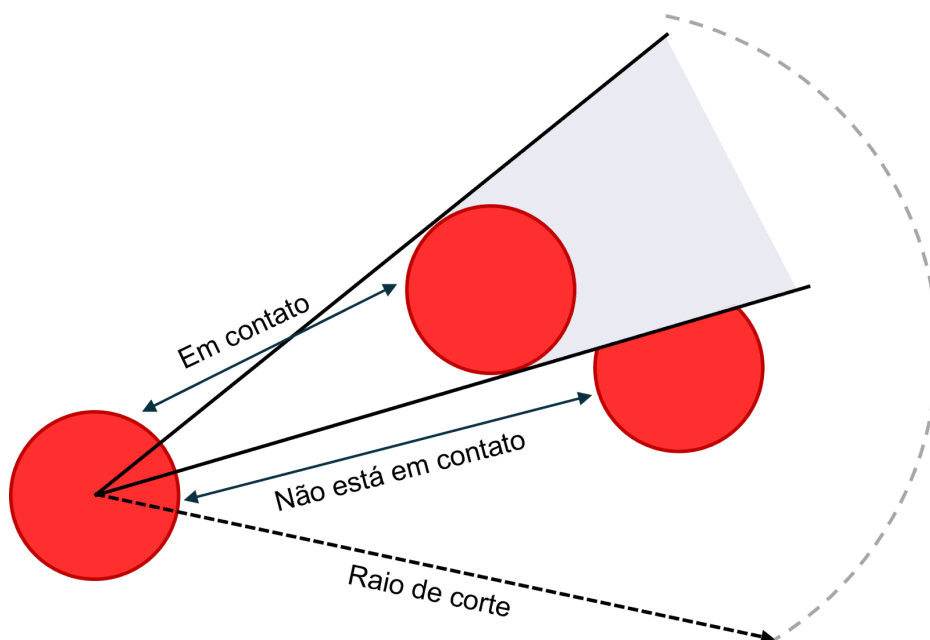
$$E_{LJ} = \sum_{\text{contatos}} \epsilon_{LJ} \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] \quad (6)$$

onde r_{ij} é a distância geométrica entre os átomos, σ_{ij} indica a posição dos contatos sendo obtido através de um mapa de contato e ϵ_{LJ} é uma constante (Oliveira; Schug; Onuchic, 2008).

Essa expressão representa o termo *contatos* da equação 1 e mostra a interação entre os Cα não ligados por ligações covalentes mas que estão em contato na estrutura nativa (Oliveira; Schug; Onuchic, 2008). O termo $\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}$ representa a repulsão e $\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10}$ a atração (Oliveira; Schug; Onuchic, 2008).

Para aplicar esse potencial, e os potenciais descritos nas seções seguintes, de forma eficaz é utilizado um mapa de contatos, que define quais pares de átomos realmente interagem na proteína, proporcionando uma representação mais detalhada e precisa da influência estrutural entre os elementos envolvidos. No presente trabalho, foi utilizado o *shadow contacts map*, que integra um critério de sombreado. Nesse método, um átomo só é considerado em contato com outro se, além de estar dentro do raio de corte de 6 Å, não houver nenhum átomo intermediário bloqueando a interação entre os dois, como exemplificado na figura 7 (Noel et al., 2012).

Figura 7 - Esquema ilustrativo do *shadow contacts map*



Fonte: Elaboração própria

Dessa forma, o *shadow contacts map* é particularmente útil em modelos baseados em estrutura, como os empregados neste estudo, por representar interações com maior precisão geométrica (Noel et al., 2012).

Entretanto, uma limitação do potencial de Lennard-Jones é que o termo de volume excluído depende da distância nativa de cada contato. Em modelos *coarse-grained*, onde as interações são definidas entre átomos de Ca que podem estar separados por até 14 Å, isso pode levar a efeitos artificiais: o tamanho efetivo das partículas em contato cresce junto com a distância nativa, fazendo com que pares distantes interajam de forma não realista em estados desenovelados (Noel et al., 2012). Embora o potencial de Lennard-Jones modificado seja adequado para contatos nativos simples, ele apresenta limitações em sistemas com múltiplos estados estruturais. Para superar essas limitações, emprega-se o potencial gaussiano, descrito a seguir.

2.1.2 Potencial Gaussiano

Diferentemente do potencial de Lennard-Jones, o potencial gaussiano permite que o volume excluído permaneça fixo independentemente da distância de contato (Retamal-Farfán et al., 2023). Isso acontece através de uma função atrativa baseada em uma curva Gaussiana que descreve a energia associada à interação entre os

resíduos nativos (Retamal-Farfán et al., 2023). Esse potencial pode ser expresso da forma

$$C_G(r_{ij}, r_0^{ij}) = \left\{ \left[1 + \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \right] [1 + G(r_{ij}, r_0^{ij})] - 1 \right\} \quad (7)$$

em que $G(r_{ij}, r_0^{ij}) = -\exp\left[-\frac{(r_{ij}-r_0^{ij})^2}{2\sigma^2}\right]$ e σ é a largura da gaussiana (Retamal-Farfán et al., 2023).

A principal vantagem desse formalismo é separar o volume excluído da posição do mínimo de energia, superando a limitação do potencial de Lennard-Jones e proporcionando uma descrição mais consistente de contatos nativos com diferentes distâncias de equilíbrio. Essa característica é essencial em sistemas que apresentam múltiplos estados estruturais estáveis, pois permite representar interações com maior fidelidade, constituindo a base para a implementação dos chamados modelos *dual basin* (Retamal-Farfán et al., 2023).

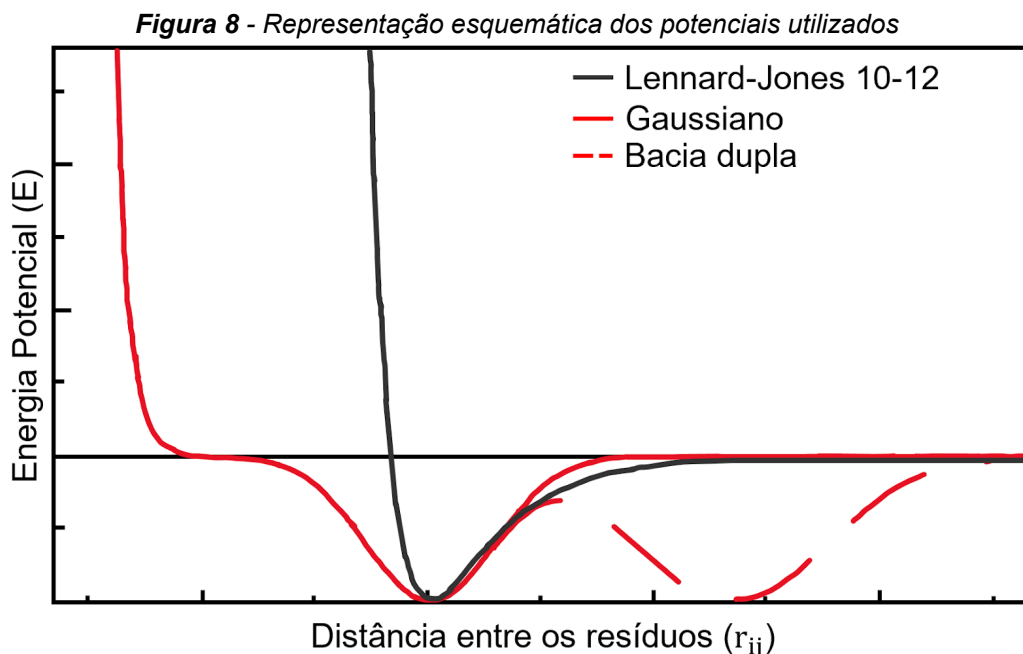
2.1.3 Potencial com duplo mínimo

No caso de interesse, em que a proteína exibe dois estados nativos, o potencial baseado em estrutura, em geral, exibe dois mínimos de energia, que podem ser implementados por meio dos chamados modelos *dual basin*. Esses modelos permitem representar simultaneamente os contatos característicos de cada conformação em uma única Hamiltoniana, possibilitando a descrição do processo de mudança conformacional entre estados nativos distintos (Retamal-Farfán et al., 2023).

A construção desse tipo de modelo envolve a junção dos mapas de contato obtidos a partir das estruturas de referência. Os contatos exclusivos de cada estado são tratados com potenciais gaussianos de um único poço, enquanto aqueles compartilhados podem assumir duas situações: se a distância nativa entre resíduos for semelhante, mantém-se apenas um termo; caso contrário, emprega-se um potencial gaussiano de dois poços, com mínimos centrados nas distâncias e equilíbrio de cada conformação (Retamal-Farfán et al., 2023). A forma geral desse potencial pode se escrita como

$$G_{DB}(r_{ij}, r_A^{ij}, r_B^{ij}) = \left\{ \left[1 + \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \right] [1 + G(r_{ij}, r_A^{ij})][1 + G(r_{ij}, r_B^{ij})] - 1 \right\} \quad (8)$$

A figura 8 apresenta uma representação esquemática dos potenciais utilizados, evidenciando a diferença entre o modelo gaussiano de um poço e o modelo *dual basin*.



Fonte: Adaptado de Noel (2012, p. 41)

Os potenciais gaussianos e, em especial, o modelo *dual basin*, permitem representar interações compatíveis com múltiplos estados nativos e simular as mudanças conformacionais entre eles. Para analisar essas transições, entretanto, não basta apenas definir o potencial: é necessário adotar métricas capazes de quantificar o progresso estrutural ao longo da simulação. Nesse sentido, as coordenadas de reação são fundamentais, pois permitem acompanhar a similaridade entre as conformações amostradas e os estados de referência da proteína.

2.2 Coordenadas de reação

Em física e química, a coordenada de reação é uma variável que indica o progresso de um processo ao longo de um caminho definido no espaço de energia potencial. Em reações químicas simples, por exemplo, ela descreve a transformação gradual de reagentes em produtos, permitindo identificar estados intermediários no

diagrama de energia da reação. De forma geral, coordenadas de reação resumem a evolução de sistemas complexos, oferecendo uma forma compacta de acompanhar o avanço de processos que envolveriam muitas variáveis simultâneas (Nelson; Cox, 2014).

Em estudos de dinâmica molecular aplicados ao enovelamento de proteínas, é comum o uso de coordenadas de reação para descrever o progresso da transição conformacional ao longo do tempo ou em função da temperatura (Cho; Levy; Wolynes, 2006). Essas coordenadas permitem acompanhar a transformação do sistema entre diferentes estados conformacionais, como os estados enovelado e desenovelado, reduzindo a complexidade multidimensional do sistema a um número limitado de variáveis relevantes (Cho; Levy; Wolynes, 2006).

O número de contatos nativos, $Q_{contato}$, mede a fração de contatos presentes em uma determinada conformação em relação à estrutura nativa (Cho; Levy; Wolynes, 2006). Para cada par de resíduos (i, j) que forma um contato nativo na estrutura de referência, verifica-se se esse contato está presente na configuração atual. A soma desses contatos, normalizada pelo total de contatos nativos, define o valor de $Q_{contato}$. Essa coordenada varia de 0 a 1, sendo 1 associado a uma conformação completamente enovelada e 0 a uma completamente desenovelada.

Outra coordenada de reação importante é a RMSD, do inglês *Root Mean Square Deviation*, que calcula o desvio médio quadrático das distâncias entre átomos equivalentes da conformação analisada e da estrutura de referência (Kufareva; Abagyan, 2012). Valores baixos dessa coordenada indicam conformações próximas ao estado nativo, enquanto valores altos estão associados a estados desenovelados. A equação da RMSD pode ser expressa como:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[(x_f - x_i)^2 + (y_f - y_i)^2 + (z_f - z_i)^2 \right]} \quad (9)$$

onde, i e f são as posições inicial e final, respectivamente, da conformação da proteína, x , y e z são as coordenadas dos resíduos e N é o número total de resíduos presentes no sistema (Kufareva; Abagyan, 2012). Embora seja simples de interpretar, o RMSD pode ser sensível a flutuações locais e nem sempre distingue bem estados intermediários de enovelamento (Kufareva; Abagyan, 2012).

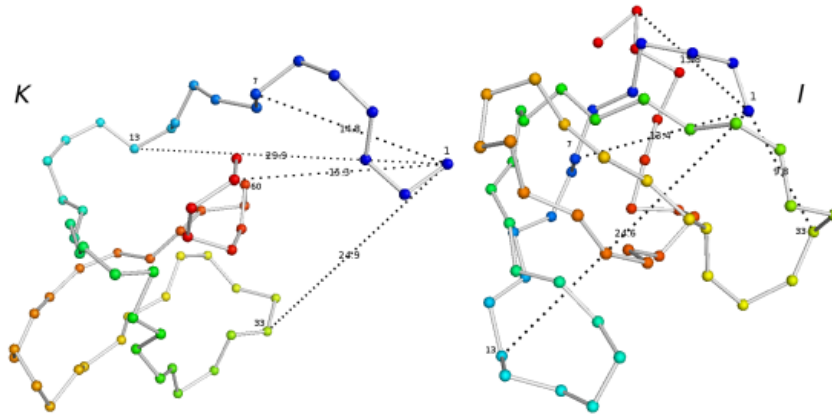
Além dessas coordenadas clássicas, outras métricas mais refinadas podem ser utilizadas para caracterizar com maior resolução as transições conformacionais. Uma delas é o q_w , que compara as distâncias internas entre os carbonos alfa de duas conformações distintas e estima o grau de similaridade estrutural entre elas (Viegas et al., 2024). Esse parâmetro é calculado a partir da seguinte equação:

$$q_w^{k,l} = \frac{1}{N} \sum_{i,j \in \text{pairs}} \exp\left(-\frac{(r_{ij}^k - r_{ij}^l)^2}{2\sigma_{ij}^2}\right) \quad (10)$$

Nessa equação, r_{ij}^k e r_{ij}^l representam as distâncias entre os resíduos i e j nas conformações k e l , respectivamente. A constante N corresponde ao número total de pares de resíduos, e $\sigma_{ij} = \sigma_0 |i - j|^\epsilon$ define a resolução de similaridade entre as distâncias internas dos pares de resíduos, atuando como um fator de tolerância, ou seja, ele determina o quanto de variação entre duas distâncias r_{ij}^k e r_{ij}^l ainda será considerado como “semelhança” dentro da função exponencial (Viegas et al., 2024).

A figura 9 ilustra um exemplo do cálculo das distâncias internas entre dois estados conformacionais distintos, identificados como k e l .

Figura 9 - Exemplo de duas conformações (k e l) nas quais são mostradas as distâncias internas entre os resíduos 1, 7, 13, 33 e 60



Fonte: Retirado de Sanches (2021, p. 18)

Enquanto métricas como $Q_{contato}$, q_w e RMSD descrevem diretamente a estrutura da proteína, também é possível caracterizar o processo de enovelamento

por meio de grandezas termodinâmicas. Um exemplo é o calor específico, cuja análise em função da temperatura permite identificar a ocorrência de transições conformacionais.

2.2.1 Calor específico

O calor específico, C_v , é uma grandeza termodinâmica que mede como a energia interna de um sistema varia com a temperatura. Em termos físicos, quando se utiliza o ensemble canônico, o sistema está em equilíbrio térmico com um reservatório que mantém a temperatura constante. O calor específico a volume constante é então definido como a derivada da energia média em relação à temperatura do reservatório:

$$C_v = \frac{d\langle E \rangle}{dT} \quad (11)$$

em que E é a energia interna e T a temperatura (Pathria, 2011).

Nas simulações de dinâmica molecular, é comum trabalhar com o *ensemble* canônico. Nesse caso, o calor específico pode ser expresso em termos das flutuações da energia interna ao longo da simulação:

$$C_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2} \quad (12)$$

onde $\langle E \rangle$ é o valor médio da energia, $\langle E^2 \rangle$ o valor médio do quadrado da energia e k_B é a constante de Boltzmann (Pathria, 2011).

A equação 12 mostra que o calor específico está diretamente relacionado à variância da energia durante a simulação, ou seja, quanto maiores as flutuações energéticas, maior o valor de C_v .

No estudo de proteínas, conforme discutido por Clementi et al. (2000), a análise do gráfico do calor específico em função da temperatura também é uma das formas de identificar transições conformacionais. O surgimento de um pico em C_v indica a temperatura na qual ocorrem maiores flutuações de energia, caracterizando uma região de alta instabilidade estrutural, indicando que o sistema transita entre

diferentes estados conformacionais. Esse pico é interpretado como a temperatura de transição conformacional ou a temperatura de enovelamento da proteína, T_f .

Apesar da utilidade das coordenadas de reação e do calor específico na caracterização de mudanças conformacionais, essas abordagens ainda dependem de estruturas de referência ou de hipóteses sobre o estado nativo da proteína. Para contornar essas limitações e obter uma representação mais abrangente do comportamento conformacional do sistema, são utilizados métodos de redução dimensional que permitem mapear o relevo de energia em espaços mais simples de interpretar. Entre esses métodos, destaca-se o ELViM, que permite mapear as diferentes conformações da proteína com base nas relações estruturais entre elas.

2.3 Energy Landscape Visualization Method

O ELViM é um método de visualização do relevo de energia das proteínas através de métodos de redução multidimensionais que utiliza como base as distâncias internas entre pares de conformações estruturais de todo o conjunto de dados analisado, não dependendo de estruturas de referência ou quaisquer outras coordenadas de reação (Viegas et al., 2024). Através de um processo iterativo, ele projeta um conjunto de conformações em um espaço bidimensional, o que permite uma análise visual clara da organização conformacional da proteína (Viegas et al., 2024).

O processo é dividido em duas etapas principais. A primeira consiste no cálculo de uma matriz de dissimilaridade que estima as distâncias estruturais entre pares de conformações em um conjunto de dados, através da métrica de q_w (equação 10) (Viegas et al., 2024).

A segunda parte do método consiste na projeção multidimensional, chamada *Force Scheme*. Nessa etapa, as conformações são representadas como pontos em um espaço bidimensional, e a posição de cada ponto é ajustada de forma a preservar as dissimilaridades calculadas (Viegas et al., 2024). O objetivo é minimizar a função de erro definida pela equação:

$$E = \sum_{(k,l)} |\delta_{k,l} - d_{k,l}| \quad (13)$$

em que $\delta_{k,l}$ representa a dissimilaridade entre as conformações k e l calculada como:

$$\delta_{k,l} = 1 - q_w^{k,l} \quad (14)$$

em que $0 < \delta_{k,l} < 1$, e $d_{k,l}$ corresponde à distância euclidiana entre os pontos k e l no plano bidimensional (Viegas et al., 2024). Dessa forma, quanto mais semelhantes estruturalmente duas conformações forem, mais próximas estarão no espaço projetado.

Além disso, é possível associar regiões distintas da projeção a diferentes estados conformacionais, o que complementa as análises feitas por meio das coordenadas de reação. Esse tipo de projeção também se destaca por permitir uma visualização direta da conectividade entre estados conformacionais, uma informação que não é facilmente acessível por meio dos gráficos de calor específico ou das coordenadas de reação isoladamente.

2.4 Detalhes computacionais

Com os potenciais e coordenadas de reação definidos, realizamos as simulações computacionais para os sistemas SH3 e KaiB, conforme detalhado a seguir.

As simulações foram realizadas com SMOG 2.4, utilizando a extensão OpenSMOG e o pacote de dinâmica molecular OpenMM. Os modelos *coarse-grained* de carbono alfa foram construídos a partir das estruturas cristalográficas disponíveis no *Protein Data Bank* (PDB): o domínio SH3 (PDB id: 1fmk) e a proteína KaiB em suas conformações gs (PDB id: 1vgl) e fs (PDB id: 5jyt).

A preparação dos arquivos envolveu a geração de um arquivo .gro com as coordenadas da estrutura *coarse-grained* e de um arquivo .top contendo os parâmetros de interação definidos pelo modelo baseado em estrutura. Para a KaiB, as simulações foram realizadas separadamente para as conformações gsKaiB e fsKaiB, utilizando o potencial de Lennard-Jones modificado para descrever os contatos nativos. Para a SH3, utilizou-se a mesma representação *coarse-grained* centrada nos átomos C α .

As simulações foram conduzidas no *ensemble* canônico, no qual o número de partículas e o volume do sistema permanecem constantes, enquanto a temperatura

é controlada por meio de um termostato Langevin. Cada simulação compreendeu 10^8 passos de integração, com trajetórias salvas a cada 10^3 passos em formato .dcd, acompanhadas de arquivos de saída com a energia total e suas componentes. As faixas de temperatura simuladas foram definidas para cada sistema, como detalhado na Seção 3.

2.5 Análise das trajetórias

As trajetórias geradas foram analisadas utilizando o pacote MDTraj e *scripts* desenvolvidos em Python, com algumas rotinas paralelizadas via Numba para maior eficiência computacional. Foram calculadas as coordenadas de reação discutidas anteriormente, incluindo o número de contatos nativos, $Q_{contato}$; o desvio médio quadrático, RMSD; e a métrica q_w . Além das métricas estruturais, foi calculado o calor específico a partir das flutuações da energia interna de cada simulação.

Os resultados foram organizados em arquivos de saída e representados graficamente, o que permitiu acompanhar a evolução conformacional da proteína em diferentes temperaturas. Por fim, os dados gerados também foram utilizados como entrada para o método ELViM, que projetou as conformações em um espaço bidimensional, facilitando a análise do relevo de energia da proteína e permitindo identificar regiões conformacionais distintas e a conectividade entre elas.

3. RESULTADOS

3.1 Análise do domínio SH3

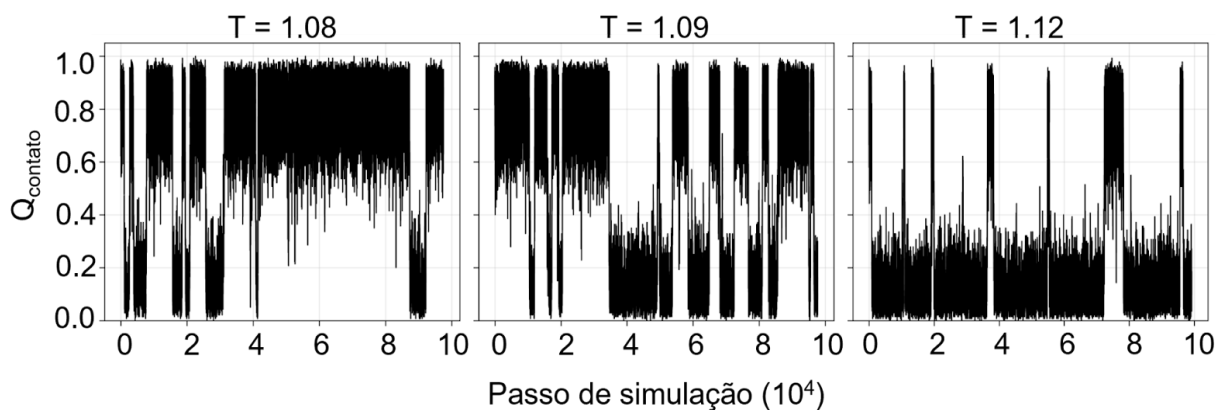
Como etapa inicial do trabalho, foram realizadas simulações com o domínio SH3, utilizado para fins de familiarização com as ferramentas computacionais e análise de enovelamento. A SH3 foi escolhida por apresentar uma única conformação nativa bem definida e um comportamento conformacional já bastante caracterizado na literatura.

Inicialmente, foram realizadas simulações para um intervalo de temperaturas variando de 0.70 a 1.50, com o objetivo de observar o comportamento geral da proteína. Para cada temperatura, foram gerados três arquivos principais: um contendo as energias do sistema, outro com os valores de força e um terceiro com a trajetória conformacional da proteína ao longo do tempo. A partir desses dados, foram construídos gráficos das coordenadas de reação $Q_{contato}$, q_w e RMSD, permitindo avaliar a estabilidade da estrutura e identificar a temperatura de transição entre os estados enovelado e desenovelado.

Com a análise desses dados, foi possível observar que, em torno da temperatura 1.09, a SH3 apresenta um comportamento em que passa parte do tempo enovelada e parte desenovelada.

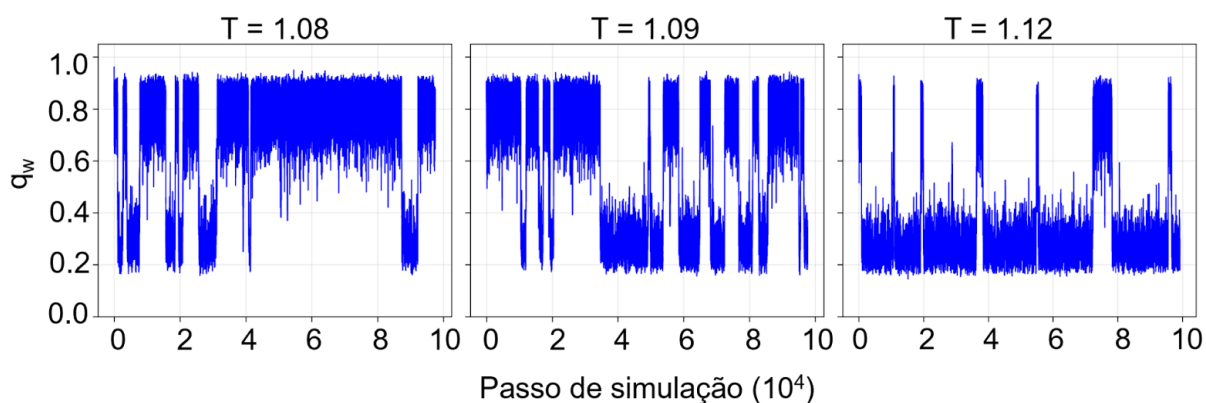
A fim de observar melhor o comportamento da proteína em torno da temperatura de enovelamento, foram analisadas, separadamente, as coordenadas de reação em três temperaturas distintas: uma mais baixa, igual a 1.08; uma na transição, igual a 1.09; e uma mais alta, igual a 1.12. As figuras 10, 11 e 12 apresentam, respectivamente, as coordenadas de reação $Q_{contato}$, q_w e RMSD ao longo da simulação para essas três temperaturas.

Figura 10 - Análise da coordenada de reação Q_{contato} ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente



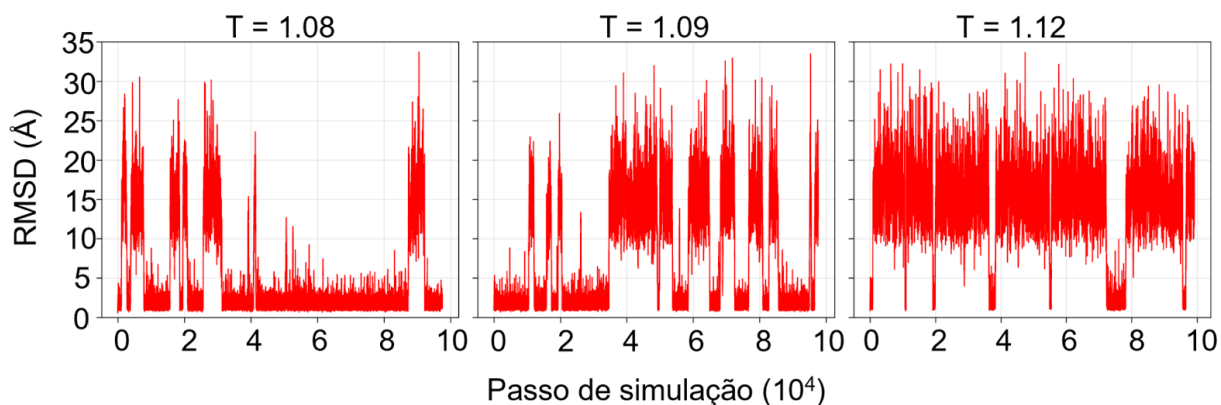
Fonte: Elaboração própria

Figura 11 - Análise da coordenada de reação q_w ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente



Fonte: Elaboração própria

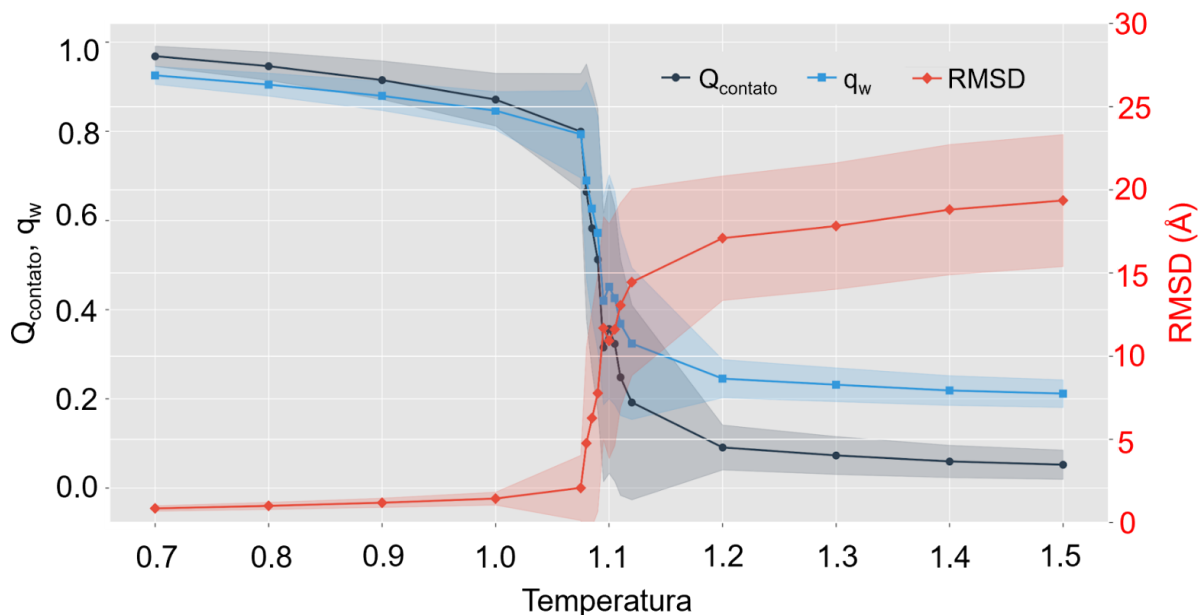
Figura 12 - Análise da coordenada de reação RMSD ao longo da simulação nas temperaturas 1.08, 1.09 e 1.12, respectivamente



Fonte: Elaboração própria

A figura 13 mostra as médias e os desvios padrão das coordenadas de reação para cada uma das simulações realizadas em diferentes temperaturas, permitindo comparar diretamente o comportamento médio de $Q_{contato}$, q_w e RMSD nas diferentes temperaturas simuladas.

Figura 13 - Gráfico das médias das coordenadas de reação em função da temperatura para a SH3



Fonte: Elaboração própria

É possível observar que em temperaturas mais baixas, os valores de $Q_{contato}$ e q_w permanecem elevados, próximos de 1, enquanto o RMSD mantém-se baixo. Esse comportamento indica que a proteína se encontra, majoritariamente, no estado nativo, com poucos desvios estruturais significativos ao longo da simulação.

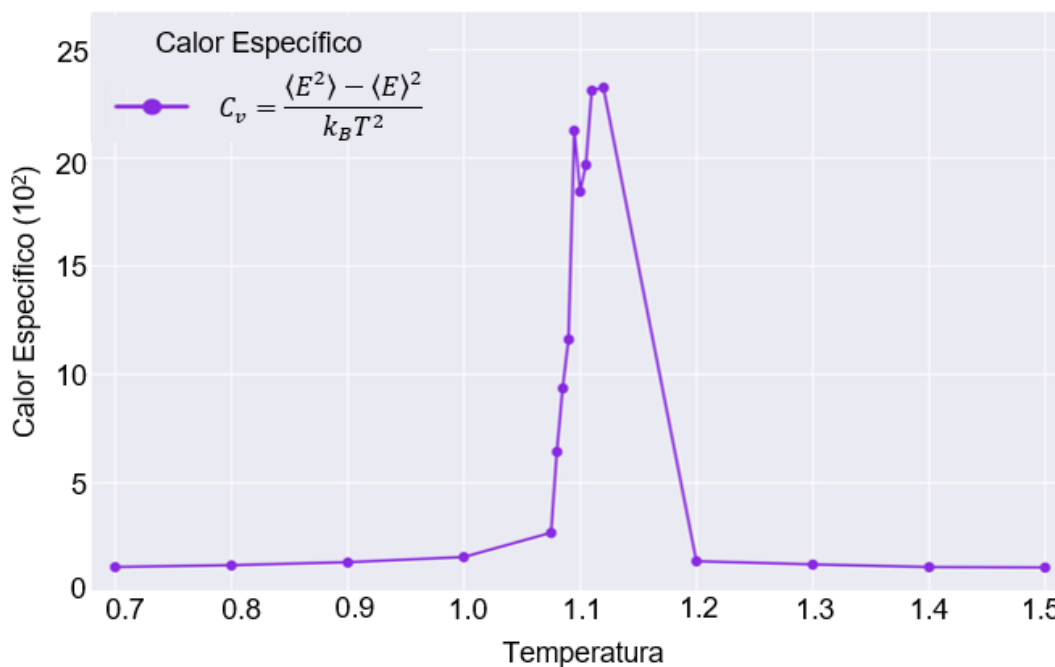
À medida que a temperatura se aproxima de 1.09, observa-se uma transição acentuada em todas as coordenadas: os valores médios de $Q_{contato}$ e q_w caem, enquanto os valores do RMSD aumentam. Essa mudança indica que a proteína está passando diretamente de um estado enovelado para um estado desenovelado. A presença de grandes desvios padrão nessa faixa de temperatura reforça a ocorrência de flutuações estruturais mais intensas, sugerindo que, durante a transição, a proteína explora várias conformações, o que reflete na dispersão dos valores em torno das médias.

É importante perceber que, embora $Q_{contato}$ e q_w descrevam tendências semelhantes, a curva de q_w apresenta uma variação mais suave. Isso acontece porque essa coordenada utiliza uma função contínua baseada nas distâncias entre os resíduos, tornando sua variação mais gradual. Já o RMSD, sendo uma medida global de similaridade estrutural em relação à estrutura nativa, representa grandes reorganizações, mas pode não capturar com a mesma precisão os detalhes locais do enovelamento.

Acima de 1.50, todas as coordenadas de reação se estabilizam em valores consistentes com o estado desenovelado: $Q_{contato}$ e q_w mantêm-se baixos, enquanto o RMSD permanece elevado, indicando que a proteína permanece fora do seu estado nativo durante o restante da simulação. Quando analisados em conjunto, esses resultados permitem estimar a temperatura de enovelamento da SH3 como sendo próxima a 1.09.

A figura 14 apresenta o calor específico da SH3 em função da temperatura. Observa-se um pico na região de $T = 1.10$, corresponde ao ponto em que a proteína apresenta as maiores flutuações energéticas, associadas à transição cooperativa entre os estados enovelado e desenovelado.

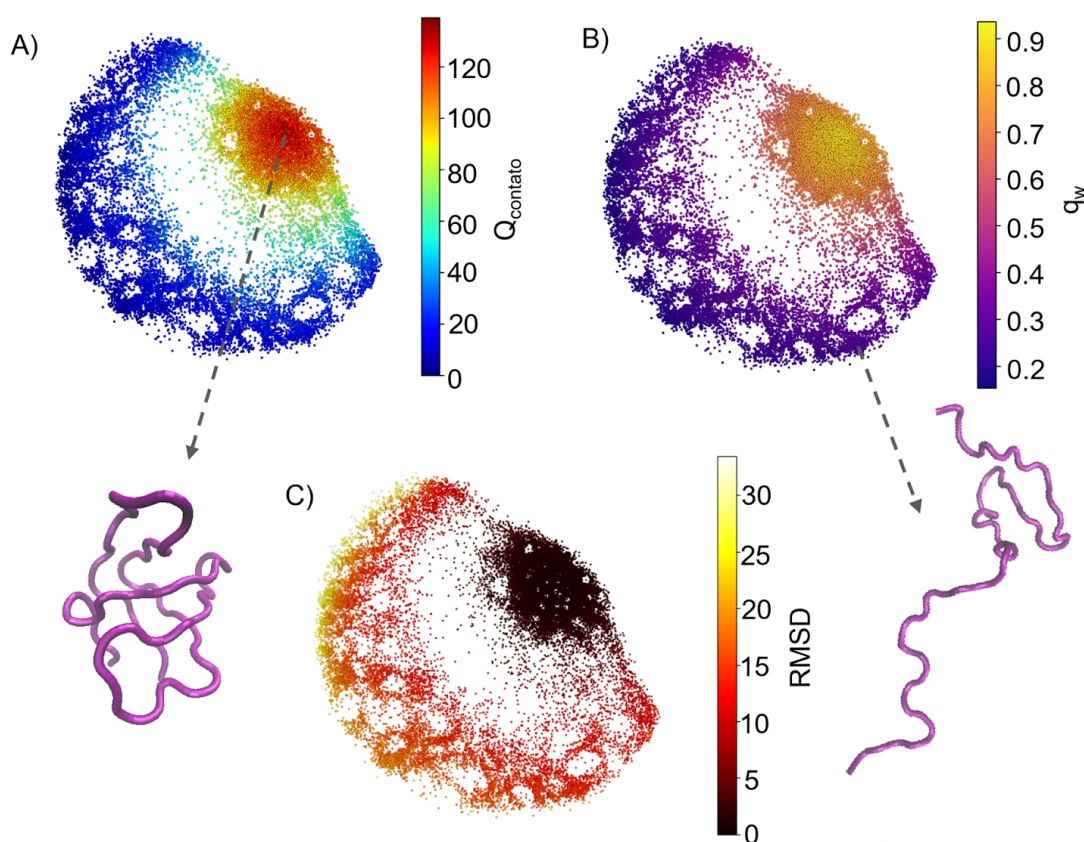
Figura 14 - Variação do calor específico da SH3 em função da temperatura



Fonte: Elaboração própria

Em seguida, com a temperatura de transição determinada, foram realizadas mais três réplicas de simulações em $T = 1.09$. Essas simulações foram posteriormente utilizadas para a construção do mapa conformacional com o método ELViM (Figura 15), descrito na seção 2.3, fornecendo uma boa base conceitual para a aplicação futura do método à proteína KaiB.

Figura 15 - Projeção bidimensional do ELViM da SH3, na temperatura 1.09, como função das coordenadas de reação $Q_{contato}$ (A), q_w (B) e RMSD (C)



Fonte: Elaboração própria

A figura 15 apresenta as projeções obtidas por meio do método ELViM para a temperatura de enovelamento. Nessa representação, conformações estruturalmente semelhantes aparecem próximas umas das outras, enquanto conformações mais distintas aparecem separadas. Cada ponto no gráfico representa uma estrutura da trajetória simulada, e a coloração indica os valores de uma coordenada de reação, $Q_{contato}$, q_w e RMSD, permitindo identificar padrões conformacionais ao longo da simulação.

Na imagem colorida por $Q_{contato}$, é possível distinguir duas regiões: uma com valores mais altos, associada a conformações próximas do estado nativo, e outra com valores mais baixos, representando estruturas desenoveladas. O mapa colorido por q_w apresenta um padrão semelhante, com regiões de maior valor indicando conformações enoveladas e valores mais baixos associados a estruturas desenoveladas.

Na projeção baseada em RMSD, também é possível identificar regiões distintas no espaço conformacional. As áreas em vermelho escuro representam conformações com baixo desvio em relação à estrutura nativa, ou seja, estruturas mais enoveladas. As demais regiões correspondem a conformações mais afastadas da estrutura nativa, indicando estados desenovelados.

Estabelecida a metodologia de análise com a SH3, a etapa seguinte consiste na investigação da proteína KaiB, considerando separadamente seus estados nativos devido à complexidade conformacional envolvida.

3.2 Análise da proteína KaiB

Como continuidade ao estudo realizado com o domínio SH3, foram conduzidas simulações para investigar separadamente o comportamento conformacional das duas conformações distintas da proteína KaiB: a gsKaiB e a fsKaiB.

A escolha por simular separadamente cada conformação teve como objetivo analisar individualmente a estabilidade térmica de cada uma delas e investigar suas respectivas temperaturas de enovelamento antes da aplicação do modelo *dual basin*. Essa análise é importante para compreender como variações na estrutura nativa influenciam a resposta conformacional ao aumento de temperatura, um aspecto central em processos como troca conformacional e regulação funcional mediada por enovelamento.

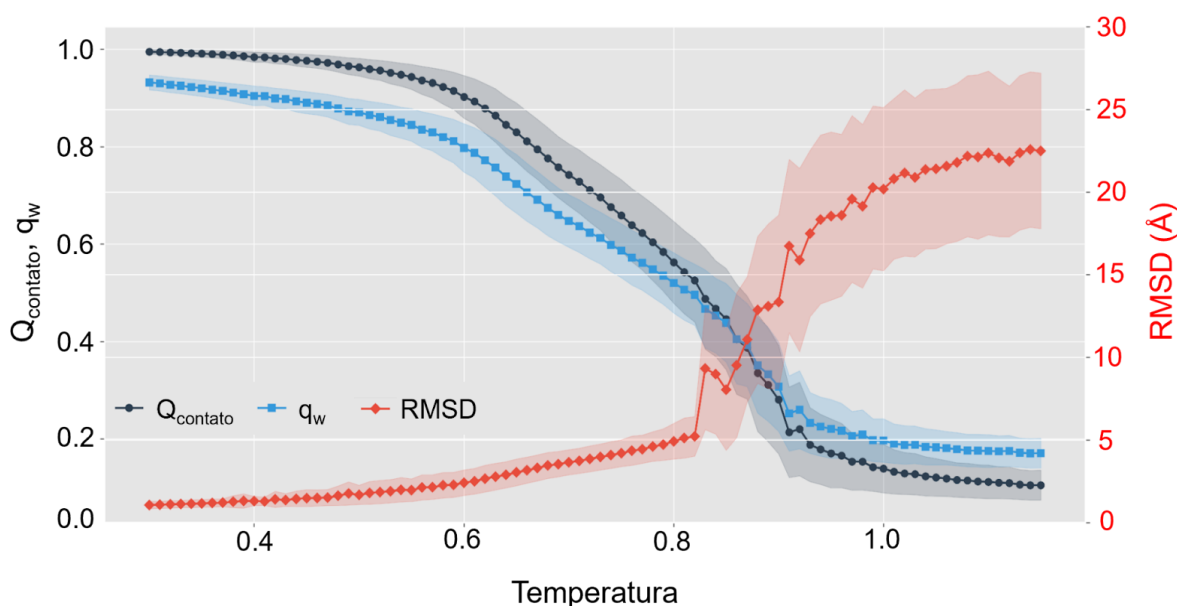
Para isso, foi realizado um conjunto de simulações cobrindo um intervalo amplo de temperaturas. No caso da gsKaiB, as temperaturas variaram de 0.30 a 1.15, enquanto para a fsKaiB foram utilizadas temperaturas entre 0.37 e 1.22. A partir das trajetórias obtidas, foram calculadas as coordenadas de reação $Q_{contato}$, q_w e RMSD ao longo do tempo. Para cada temperatura simulada foi determinada a média dessas coordenadas, o que possibilitou construir gráficos que ilustram como

cada estrutura responde ao aumento de temperatura e em que ponto ocorre sua transição conformacional.

3.2.1 Conformação gsKaiB

No gráfico da gsKaiB (Figura 16), observa-se que em temperaturas mais baixas os valores médios de $Q_{contato}$ e q_w se mantêm próximos de 1, enquanto o RMSD permanece reduzido, indicando pouca variação estrutural em relação à conformação de referência. Esse comportamento, típico de um estado enovelado, sugere que a conformação gsKaiB permanece predominantemente enovelada e estável nessas condições.

Figura 16 - Gráfico das médias das coordenadas de reação em função da temperatura para a conformação gsKaiB



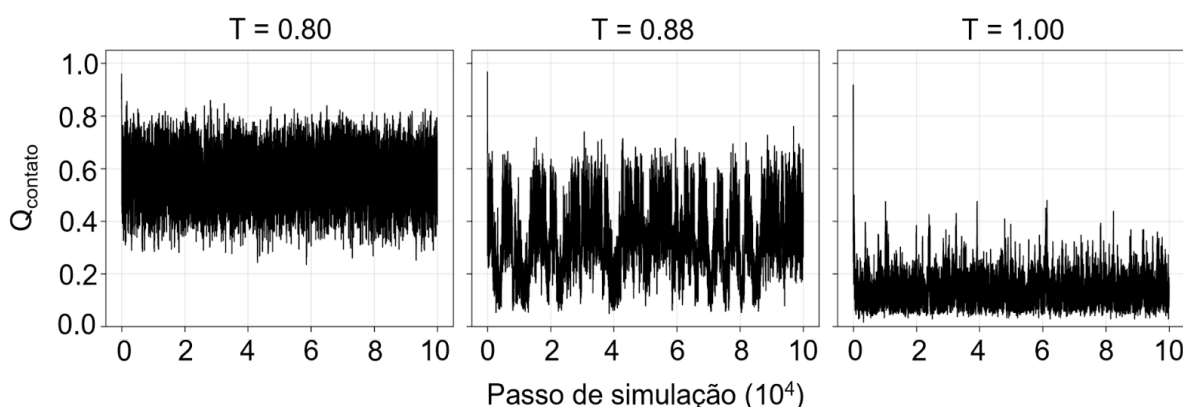
Fonte: Elaboração própria

À medida que a temperatura aumenta, nota-se uma transição progressiva nas coordenadas de reação: $Q_{contato}$ e q_w apresentam uma queda gradual, enquanto o RMSD cresce de forma contínua. Essa mudança indica que a proteína está deixando de ocupar preferencialmente estados enovelados e passa a explorar estados cada vez mais desenovelados. Essa transição, no entanto, ocorre de maneira menos abrupta do que a observada na SH3, o que pode ser atribuído a uma menor cooperatividade estrutural da gsKaiB.

A presença de desvios padrão mais acentuados em uma faixa intermediária de temperaturas sugere uma fase instável, com flutuações mais intensas, indicando que a proteína alterna entre diferentes conformações.

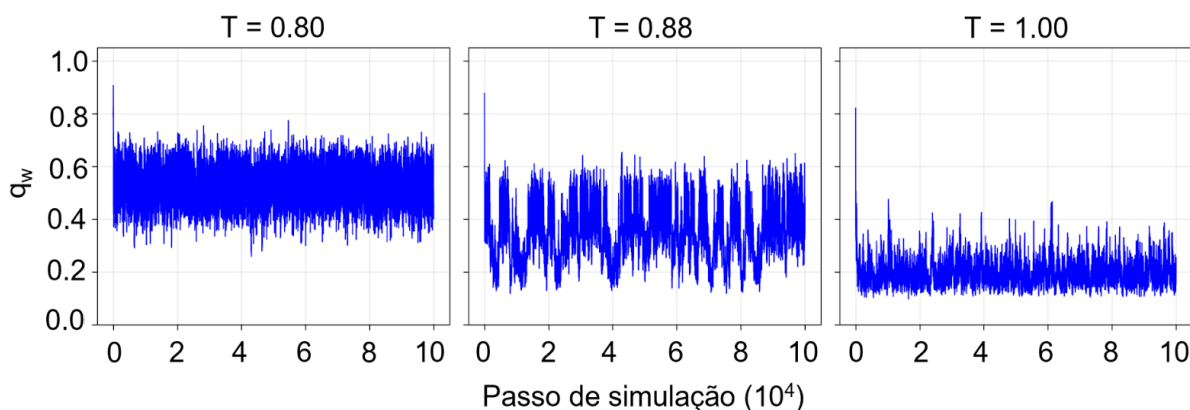
De forma complementar, as figuras 17, 18 e 19 a seguir mostram a evolução temporal de $Q_{contato}$, q_w e RMSD para três temperaturas representativas: $T = 0.80$, $T = 0.88$ e $T = 1.00$.

Figura 17 - Análise da coordenada de reação $Q_{contato}$ ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente



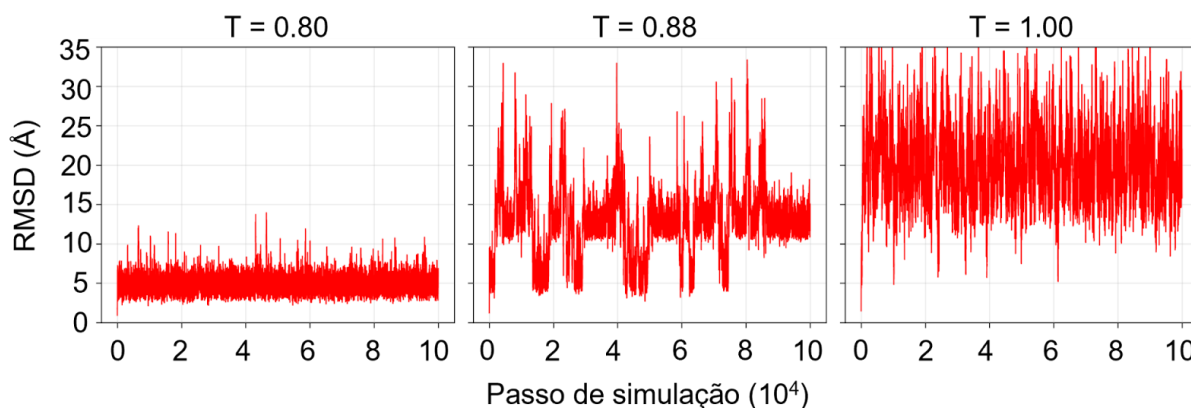
Fonte: Elaboração própria

Figura 18 - Análise da coordenada de reação q_w ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente



Fonte: Elaboração própria

Figura 19 - Análise da coordenada de reação *RMSD* ao longo da simulação nas temperaturas 0.80, 0.88 e 1.00, respectivamente

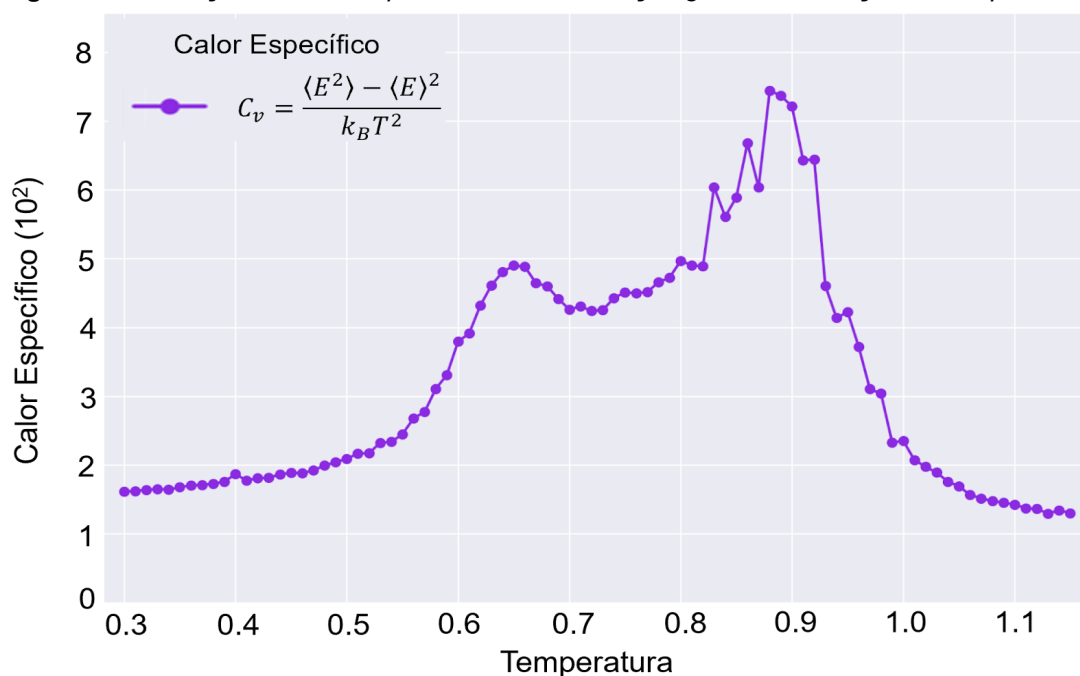


Fonte: Elaboração própria

O comportamento das curvas reforça a caracterização feita com base nas médias: em $T = 0.80$, a proteína se mantém majoritariamente enovelada, enquanto em $T = 1.00$ predominam configurações desenoveladas. A maior flutuação em $T = 0.88$ é compatível com uma região de transição conformacional.

A análise do calor específico (Figura 20) confirma essa transição estrutural: o pico em torno de $T = 0.88$ é interpretado como a temperatura de enovelamento da gsKaiB.

Figura 20 - Variação do calor específico da conformação gsKaiB em função da temperatura

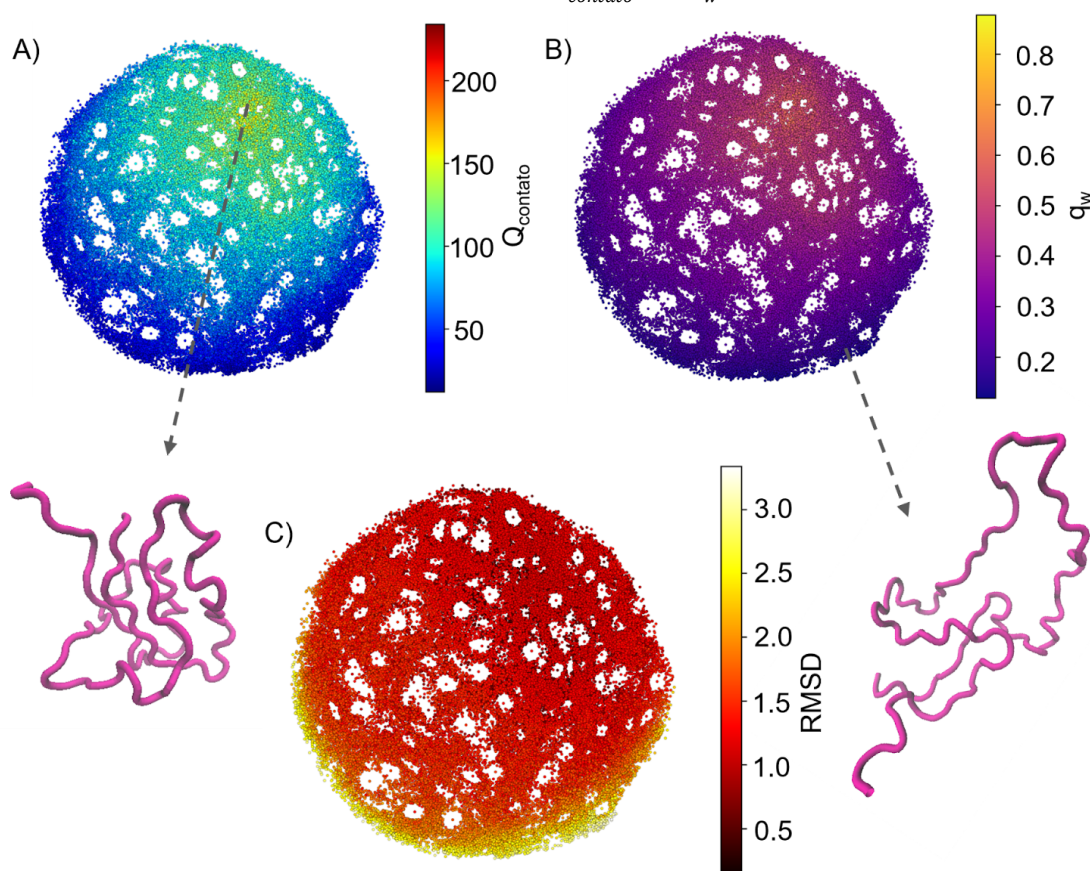


Fonte: Elaboração própria

Essa conclusão é consistente com o comportamento dinâmico observado nos gráficos das coordenadas de reação ao longo do tempo, em que se nota uma maior flutuação estrutural nessa faixa de temperatura. Acima dessa temperatura, todas as coordenadas analisadas se estabilizam em padrões consistentes com o estado desenovelado, caracterizado por baixos valores de $Q_{contato}$ e q_w e altos valores de RMSD.

A fim de investigar como as diferentes conformações acessadas pela gsKaiB se distribuem no espaço conformacional próximo à temperatura associada ao pico do calor específico, foi realizada a projeção pelo método ELViM. A figura 21 mostra a distribuição dos estados obtida para a temperatura $T = 0.88$, com a coloração representando os valores de $Q_{contato}$, q_w e RMSD em relação à estrutura de referência gsKaiB.

Figura 21 - Projeção bidimensional do ELViM da gsKaiB, na temperatura 0.88, como função das coordenadas de reação $Q_{contato}$ (A), q_w (B) e RMSD (C)



Fonte: Elaboração própria

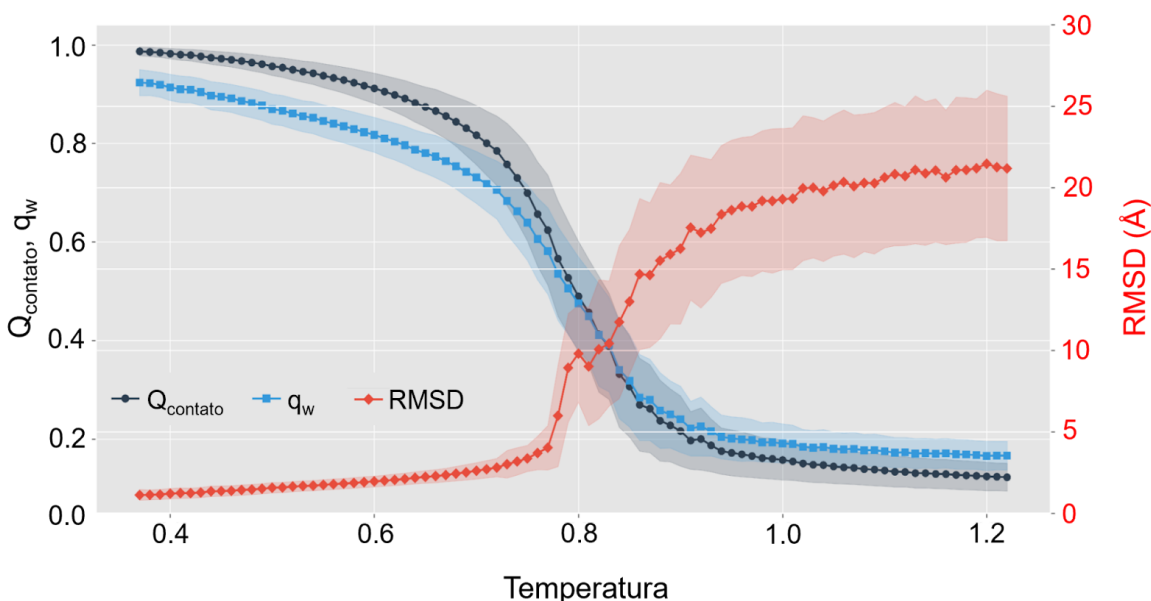
Observa-se que a projeção da gsKaiB não apresenta regiões bem delimitadas que possam ser diretamente associadas aos estados enovelado e desenovelado, o que é compatível com superfícies de energia mais rugosas, indicando múltiplas rotas de transição. Em vez disso, o espaço conformacional é preenchido de forma mais contínua, com valores intermediários de $Q_{contato}$ e q_w predominando e RMSD relativamente elevados em grande parte da trajetória. Esse padrão sugere que a transição conformacional da gsKaiB é mais distribuída, possivelmente envolvendo múltiplos estados intermediários, o que está de acordo com a presença de um pico de calor específico, mas sem evidência clara de dois estados estruturais bem definidos.

Após analisar a estabilidade térmica da gsKaiB, voltamos nossa atenção à fsKaiB, a fim de investigar se ela exibe um padrão semelhante ou distinto de transição conformacional.

3.2.2 Conformação fsKaiB

No caso da fsKaiB (Figura 22), observa-se um comportamento similar ao da gsKaiB em temperaturas mais baixas: $Q_{contato}$ e q_w mantêm-se elevados, próximos de 1, e o RMSD permanece reduzido, indicando predominância do estado enovelado.

Figura 22 - Gráfico das médias das coordenadas de reação em função da temperatura para a conformação fsKaiB



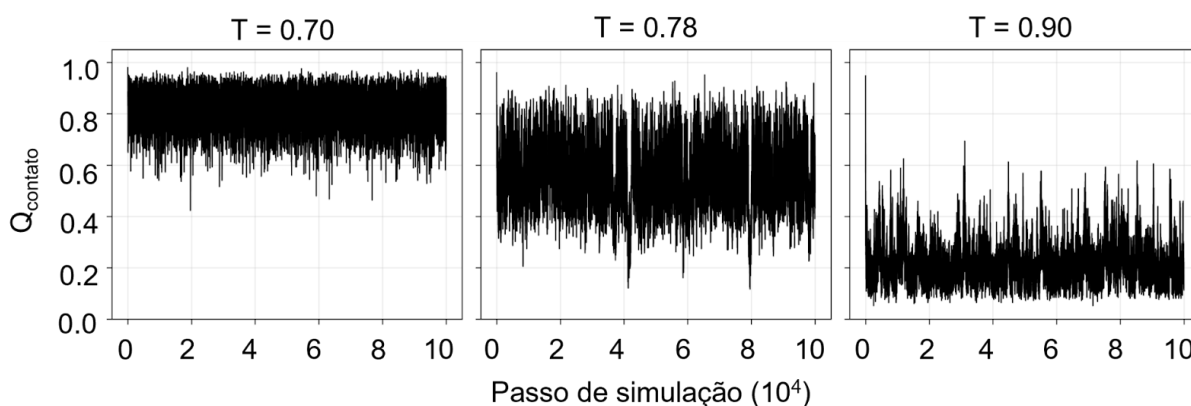
Fonte: Elaboração própria

Com o aumento da temperatura, os valores médios de $Q_{contato}$ e q_w diminuem, enquanto o RMSD aumenta. O padrão de transição indica uma mudança contínua no equilíbrio conformacional da proteína, sugerindo que a fsKaiB também apresenta uma transição térmica entre estados enovelado e desenovelado.

Assim como na gsKaiB, a região de maior variação nas coordenadas de reação coincide com um aumento nos desvios padrão, refletindo a presença de flutuações estruturais mais intensas.

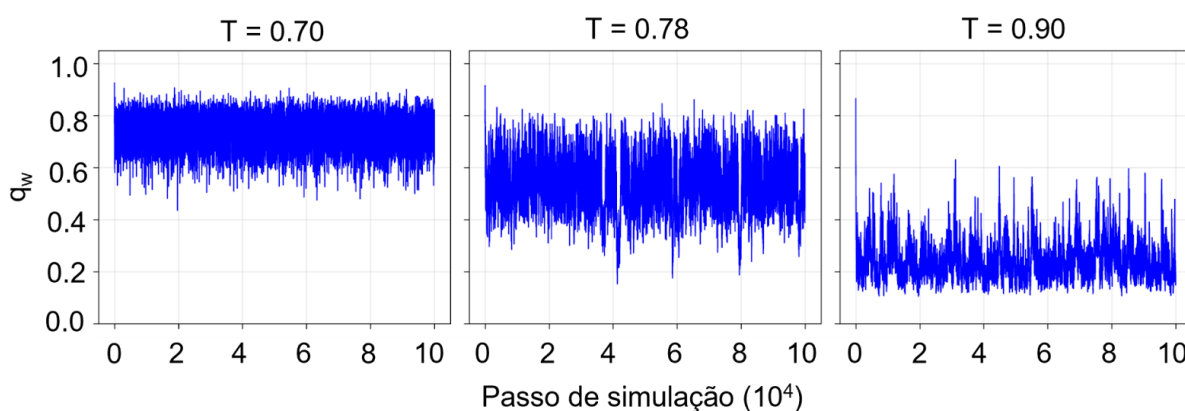
Para ilustrar qualitativamente as flutuações conformacionais ao longo do tempo, as figuras 23, 24 e 25 apresentam a variação de $Q_{contato}$, q_w e RMSD durante a simulação para três temperaturas: uma abaixo da transição, em $T = 0.70$, uma próxima à temperatura de enovelamento, $T = 0.78$, e uma acima, em $T = 0.90$.

Figura 23 - Análise da coordenada de reação $Q_{contato}$ ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente



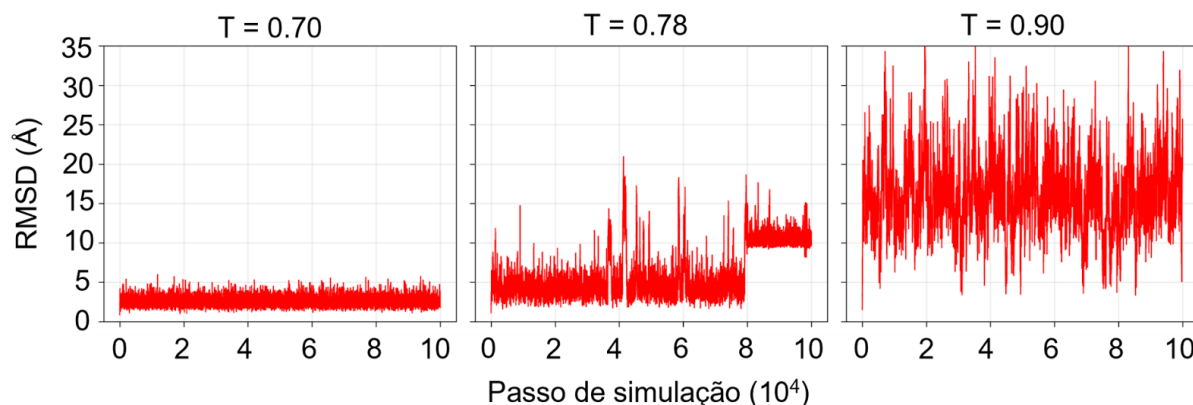
Fonte: Elaboração própria

Figura 24 - Análise da coordenada de reação q_w ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente



Fonte: Elaboração própria

Figura 25 - Análise da coordenada de reação RMSD ao longo da simulação nas temperaturas 0.70, 0.78 e 0.90, respectivamente

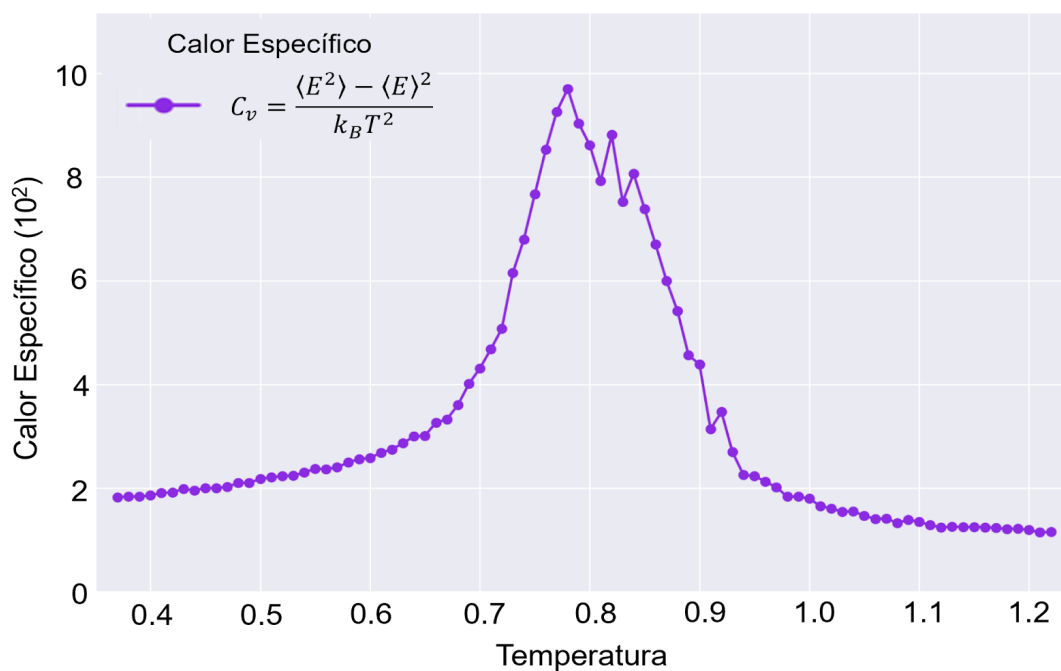


Fonte: Elaboração própria

Observa-se que, nas temperaturas mais baixas, os valores das coordenadas de reação permanecem relativamente estáveis, enquanto em $T = 0.78$ há maior variabilidade, indicando a ocorrência de transições entre diferentes conformações. Já em $T = 0.90$, os valores se mantêm baixos para $Q_{contato}$, q_w , e elevados para RMSD, consistentes com o regime desenovelado predominante.

Uma estimativa melhor para a temperatura de enovelamento foi obtida por meio da análise do calor específico (Figura 26), cujo pico ocorreu em torno de 0.78. Esse valor é, portanto, considerado como a temperatura de enovelamento da fsKaiB. Acima dessa temperatura, a proteína tende a permanecer em estados desenovelados, com baixos valores de $Q_{contato}$ e q_w e altos valores de RMSD, caracterizando o regime de desenovelamento.

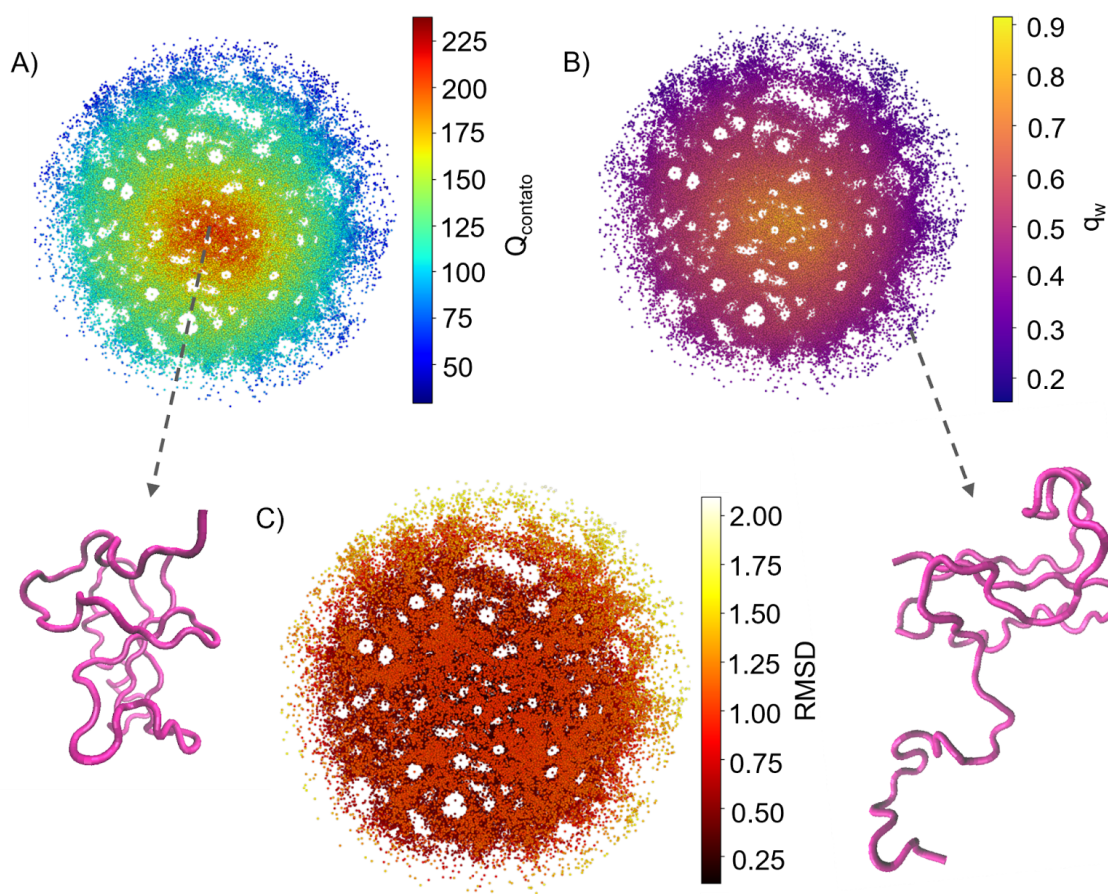
Figura 26 - Variação do calor específico da proteína fsKaiB em função da temperatura



Fonte: Elaboração própria

Com o objetivo de visualizar a paisagem conformacional explorada pela proteína fsKaiB, foi aplicada a metodologia ELViM a uma simulação realizada na temperatura $T = 0.78$. A figura 27 mostra a projeção das conformações da fsKaiB em um espaço bidimensional obtido com o método ELViM.

Figura 27 - Projeção bidimensional do ELViM da fsKaiB, na temperatura 0.78, como função das coordenadas de reação $Q_{contato}$ (A), q_w (B) e RMSD (C), respectivamente



Fonte: Elaboração própria

Os mapas apresentam uma distribuição densa e contínua de pontos, sem a presença de estados bem definidos ou separados por barreiras conformacionais visíveis, como observado no domínio SH3. Ainda assim, é possível identificar regiões centrais com valores mais altos de $Q_{contato}$ e q_w , correspondendo a estruturas mais próximas da conformação nativa. Por outro lado, as bordas da projeção apresentam valores menores dessas coordenadas, indicando conformações mais desorganizadas. O gradiente suave nas três representações sugere que, nesta temperatura, a fsKaiB se encontra em um regime de transição, com a coexistência de múltiplas conformações intermediárias, o que está de acordo com os resultados apresentados anteriormente para as médias das coordenadas de reação e o pico do calor específico.

Com base nos dados apresentados, é possível realizar uma comparação direta com a conformação gsKaiB. Nota-se que a transição da gsKaiB ocorre em

temperatura superior à da fsKaiB, indicando maior estabilidade térmica. Essa diferença pode estar relacionada a variações no número e na distribuição dos contatos nativos ou à maior compactação da estrutura gsKaiB, em concordância com os papéis funcionais distintos de cada conformação ao longo do ciclo circadiano, reforçando a importância de sua caracterização térmica separada.

4. DISCUSSÃO

A análise das simulações realizadas permitiu avaliar diferenças no comportamento conformacional entre o domínio SH3 e a KaiB nas suas duas conformações gsKaiB e fsKaiB.

A SH3 apresentou uma transição conformacional bem definida, com variações abruptas nas coordenadas de reação ao redor da temperatura de 1.09. Esse comportamento é típico de sistemas com alta cooperatividade estrutural, nos quais a quebra de contatos nativos ocorre de forma coletiva sem a presença de muitos intermediários. Conforme discutido por Clementi et al. (2000) e Oliveira et al. (2008), a cooperatividade elevada está associada a superfícies de energia com um único mínimo, favorecendo uma resposta termodinâmica rápida e bem definida. Isso reforça a adequação da SH3 como sistema modelo para testes metodológicos, dado seu padrão de enovelamento direto e estrutura compacta.

Em comparação com a SH3, tanto a gsKaiB quanto a fsKaiB apresentaram transições térmicas mais suaves, refletidas por variações graduais nas coordenadas de reação e picos de calor específico mais largos. Esse comportamento sugere uma menor cooperatividade estrutural, característica esperada em proteínas com múltiplos estados funcionais. A presença de caminhos conformacionais alternativos ou intermediários menos definidos é compatível com a função de *fold-switch* da KaiB, cuja dinâmica requer flexibilidade para alternar entre dois estados nativos distintos.

A transição térmica da gsKaiB ocorre de forma contínua, com uma mudança gradual no equilíbrio conformacional. A análise do calor específico mostra um pico em 0.88, sugerindo que essa é a temperatura de enovelamento. O fato de essa temperatura ser mais elevada que a observada para a fsKaiB pode indicar uma maior estabilidade térmica da gsKaiB, o que é coerente com seu papel como conformação basal e funcionalmente inativa da KaiB durante grande parte do ciclo circadiano, sugerindo que sua estabilidade pode ser essencial para garantir controle temporal preciso.

A fsKaiB apresenta uma queda progressiva nas coordenadas de reação e aumento do RMSD conforme a temperatura se eleva. A temperatura de transição estimada em 0.78 é mais baixa, sugerindo menor estabilidade térmica, possivelmente refletindo a natureza transitória dessa conformação. Esse comportamento pode estar relacionado à sua função como estado ativado da KaiB,

acessado em fases específicas do ciclo circadiano. A menor cooperatividade observada reforça a ideia de que sua estrutura admite um maior número de estados intermediários, necessários para a flexibilidade conformacional envolvida no mecanismo de *fold-switch*.

Por fim, a análise do calor específico complementa as observações feitas a partir das coordenadas de reação. Os picos observados em torno de 0.88 para a gsKaiB e de 0.78 para a fsKaiB indicam regiões de maior flutuação da energia, o que está de acordo com as tendências identificadas nas curvas de $Q_{contato}$, q_w e RMSD. Esses resultados fornecem uma evidência adicional para a localização das temperaturas de transição de cada conformação, o que está de acordo com a definição clássica de temperatura de enovelamento, ou seja, a temperatura em que as populações dos estados enovelado e desenovelado se igualam. A combinação dessas abordagens permitiu caracterizar com maior confiança a estabilidade térmica de cada estrutura, fornecendo uma base sólida para investigações futuras com o modelo *dual basin*, que explora explicitamente a transição entre os dois estados da KaiB.

Assim, o contraste entre a SH3 e as duas conformações da KaiB evidencia a diversidade de mecanismos conformacionais presentes nas proteínas, com implicações diretas para sua função biológica e regulação. A SH3, com seu enovelamento altamente cooperativo, serve como um modelo ideal de transição nítida, enquanto a KaiB ilustra a importância de transições graduais e flexíveis em sistemas biológicos mais complexos.

5. CONCLUSÃO

Este trabalho teve como objetivo investigar as transições conformacionais da proteína KaiB, com ênfase no fenômeno de *fold-switch*, utilizando modelos computacionais baseados em estrutura. As simulações realizadas permitiram caracterizar separadamente o comportamento térmico das duas conformações da KaiB, gsKaiB e fsKaiB, determinando suas respectivas temperaturas de enovelamento como 0.88 e 0.78.

Os resultados mostraram que ambas as conformações passam por transições entre estados enovelado e desenovelado, embora de maneira menos abrupta quando comparado ao observado para a SH3. Essa diferença sugere uma menor cooperatividade estrutural da KaiB, possivelmente relacionada à sua capacidade de alternar entre conformações distintas.

Embora o estudo com o modelo *dual basin* para a KaiB não tenha sido completamente desenvolvido devido a limitações de tempo, essa abordagem representa uma direção natural para a continuidade desta pesquisa, permitindo investigar simultaneamente as duas conformações em um único modelo energético.

De modo geral, este trabalho cumpriu seus objetivos ao utilizar a SH3 como sistema modelo para familiarização com os procedimentos e ferramentas computacionais empregados no estudo de transições conformacionais, além de realizar simulações de dinâmica molecular das conformações da KaiB em diferentes temperaturas, permitindo a identificação de suas respectivas temperaturas de enovelamento e a análise comparativa de suas estabilidades. A construção de um modelo *dual basin* para a KaiB foi iniciada, e sua aplicação futura permitirá investigar de forma mais aprofundada a transição entre as conformações gsKaiB e fsKaiB.

REFERÊNCIAS

- ALBERTS, Bruce *et al.* **Biologia Molecular Da Célula**. 6. ed. Porto Alegre: Artmed, 2017.
- ANFINSEN, C. B. Principles that govern the folding of protein chains. **Science (New York, N.Y.)**, v. 181, n. 4096, p. 223–230, 20 jul. 1973.
- CHANG, Yong-Gang *et al.* A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. **Science**, v. 349, n. 6245, p. 324–328, 17 jul. 2015.
- CHO, Samuel S.; LEVY, Yaakov; WOLYNES, Peter G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. **Proceedings of the National Academy of Sciences**, v. 103, n. 3, p. 586–591, 17 jan. 2006.
- CLEMENTI, Cecilia; NYMEYER, Hugh; ONUCHIC, José Nelson. 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**, v. 298, n. 5, p. 937–953, 19 maio 2000.
- COHEN, Susan E.; GOLDEN, Susan S. Circadian Rhythms in Cyanobacteria. **Microbiology and Molecular Biology Reviews**, v. 79, n. 4, p. 373–385, 2 set. 2015.
- CONTESSOTO, Vinícius de Godoi *et al.* Introdução ao problema de enovelamento de proteínas: uma abordagem utilizando modelos computacionais simplificados. **Revista Brasileira de Ensino de Física**, v. 40, p. e4307, 18 jun. 2018.
- DOBSON, Christopher M. Protein folding and misfolding. **Nature**, v. 426, n. 6968, p. 884–890, dez. 2003.
- GÜNTHER, Florian Steffen. **Strukturoptimierung von Proteinmodellen mittels Zyklen aus lokaler Störung und nachfolgendem Quench**. 2013. 97 f. Masterarbeit (Master of Science) – Fakultät für Naturwissenschaften, Technische Universität Chemnitz, Chemnitz, 2013.
- HUBNER, Isaac A.; EDMONDS, Katherine A.; SHAKHNOVICH, Eugene I. Nucleation and the Transition State of the SH3 Domain. **Journal of Molecular Biology**, v. 349, n. 2, p. 424–434, 3 jun. 2005.
- KUFAREVA, Irina; ABAGYAN, Ruben. Methods of protein structure comparison. **Methods in molecular biology (Clifton, N.J.)**, v. 857, p. 231–257, 2012.
- NELSON, David L.; COX, Michael M. **Princípios De Bioquímica De Lehninger**. 6. ed. [S.l.]: Artmed, 2014.
- NOEL, Jeffrey K.; ONUCHIC, José N. The Many Faces of Structure-Based Potentials: From Protein Folding Landscapes to Structural Characterization of Complex Biomolecules. *In*: DOKHOLYAN, Nikolay V. (Org.). **Computational Modeling of Biological Systems**. Biological and Medical Physics, Biomedical Engineering. Boston, MA: Springer US, 2012. p. 31–54.

NOEL, Jeffrey K.; WHITFORD, Paul C.; ONUCHIC, José N. The Shadow Map: A General Context Definition for Capturing the Dynamics of Biomolecular Folding and Formation. **The Journal of Physical Chemistry B**, v. 116, n. 29, p. 6626-6702, 28 jul. 2012.

OLIVEIRA JUNIOR, Antonio Bento de. **Métodos para visualização de superfície de energia do enovelamento de proteínas**. São José do Rio Preto: Universidade Estadual Paulista, 6 dez. 2017.

OLIVEIRA, Leandro C.; SCHUG, Alexandre; ONUCHIC, José N. Geometrical Features of the Protein Folding Mechanism Are a Robust Property of the Energy Landscape: A Detailed Investigation of Several Reduced Models. **The Journal of Physical Chemistry B**, v. 112, n. 19, p. 6131-6136, 1 maio 2009.

PATHRIA, R. K. **Statistical Mechanics**. 3. ed. [S.l.]: Academic Press, 2011.

RETAMAL-FARFÁN, Ignacio et al. Exploring the structural acrobatics of fold-switching proteins using simplified structure-based models. **Biophysical Reviews**, v. 15, n. 4, p. 787-799, 1 ago. 2023.

SANCHES, Murilo Nogueira. **Representação detalhada do relevo de energia de monômeros de β -amilóide e de proteínas de repetições**. 15 dec. 2021.

THE NOBEL PRIZE. **The Nobel Prize in Chemistry 2024**. Disponível em: <https://www.nobelprize.org/prizes/chemistry/2024/summary/>. Acesso em: 25 set. 2025.

VIEGAS, Rafael Giordano *et al.* ELViM: Exploring Biomolecular Energy Landscapes through Multidimensional Visualization. **Journal of Chemical Information and Modeling**, v. 64, n. 8, p. 3443–3450, 22 abr. 2024.