



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Tese de Doutorado

**SIMULAÇÃO MOLECULAR DE PROTEÍNAS
COM NÓS GEOMÉTRICOS**

José Cícero Alves da Silva

Orientador: Prof. Dr. Gerd Bruno Rocha

Coorientador: Dr. Gabriel Aires Urquiza de Carvalho

João Pessoa

-Maio de 2022-



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós-Graduação em Química

Tese de Doutorado

Simulação Molecular de Proteínas com Nós Geométricos

José Cícero Alves da Silva

Tese submetida ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba, como parte dos requisitos para obtenção do título de Doutor em Ciências com área de concentração em Físico-química.

Orientador: Prof. Dr. Gerd Bruno Rocha

Coorientador: Dr. Gabriel Aires Urquiza de Carvalho

João Pessoa

-Maio de 2022-

Simulação molecular de proteínas com nós geométricos.

Tese de Doutorado apresentada pelo aluno José Cícero Alves da Silva e aprovada pela banca examinadora em 30 de maio de 2022.

Gerd Bruno da Rocha

Prof. Dr. Gerd Bruno da Rocha
Departamento de Química – CCEN/UFPB
Orientador/Presidente

Gabriel Urquiza

Dr. Gabriel Aires Urquiza de Cavalho
Pesquisador junto a UFPE
2º. Orientador

Carlos Henrique da Silveira

Prof. Dr. Carlos Henrique da Silveira
ICT/UNIFEI – Campus Itabira-MG
Examinador

Sidney Ramos de Santana

Prof. Dr. Sidney Ramos de Santana
DQF/UFPB-Areia-PB
Examinador

Karen C. Weber

Profa. Dra. Karen Cacilda Weber
Departamento de Química – UFPB-PB
Examinadora

Otávio Luís de Santana

Prof. Dr. Otávio Luís de Santana
Departamento de Química – UFPB-PB
Examinador

Catálogo na publicação
Seção de Catalogação e Classificação

S586s Silva, Jose Cicero Alves da.
Simulação molecular de proteínas com nós geométricos
/ Jose Cicero Alves da Silva. - João Pessoa, 2022.
101 f. : il.

Orientação: Gerd Bruno Rocha.
Coorientação: Gabriel Aires Urquiza de Carvalho.
Tese (Doutorado) - UFPB/CCEN.

1. Físico-química. 2. PcN. 3. Dinâmica molecular. 4.
PCA. 5. DCCM. 6. Descritores de reatividade. I. Rocha,
Gerd Bruno. II. Carvalho, Gabriel Aires Urquiza de.
III. Título.

UFPB/BC

CDU 541.1(043)

Dedico esta Tese à minha 'vovó' Maria Teixeira de Melo (in memoriam), por ter me ensinado que fé, humildade e respeito ao próximo são os maiores valores que um ser humano pode e precisa ter.

*"Até cortar os nossos próprios defeitos
pode ser perigoso. Nunca se sabe qual é o
defeito que sustenta o nosso edifício inteiro."*

Clarisse Lispector

AGRADECIMENTOS

- A humanidade que se apresenta em cada pessoa e de pessoa para pessoa.
- A minha família.
- Ao professor Gerd Rocha pelo aceite de orientação ao longo de um período tão desafiador. Ao meu coorientador, Gabriel Urquiza pelas contribuições no desenvolvimento do trabalho.
- Aos meus amigos do Laboratório de Química Quântica Computacional (LQQC) Larissa, Emerson, Acassio, Elton, Igor, Júlio, Gutemberg e Eduardo por todos os momentos compartilhados.
- As professoras Cícera (primeira série), Márcia Poliana (ensino médio) e Andréa Monteiro (graduação) por mesmo quando o cenário não foi nada favorável, oferecer esperança, palavras que fizeram toda diferença.
- A minha grande amiga Geane pela amizade, carinho e partilha há tantos anos. Aos meus amigos Lorrán e Hugo pelo apoio e amizade durante essa caminhada.
- Aos professores do PPGQ/UFPB Karen Weber, Oscar Malta e Otávio Santana pelo aprendizado e exemplos a serem seguidos.
- À banca examinadora: Prof. Carlos Silveira, Prof. Sidney Ramos, Profa. Karen Weber, Prof. Otávio Santana, Prof. Ítalo Curvelo e Prof. Wagner Faustino pelo aceite e todas as contribuições que certamente enriqueceram o trabalho.
- A Marcos Pequeno, pela presteza em ajudar frente às demandas na secretaria do PPGQ/UFPB.
- À Universidade Federal da Paraíba e ao Programa de Pós-graduação em Química (PPGQ/UFPB).
- As agências de amparo CNPq e CAPES pelo suporte financeiro. Ao CENAPAD-SP pelo tempo computacional concedido.

RESUMO

O papel dos nós nas proteínas permanece indefinido. Alguns estudos sugerem um impacto na estabilidade, no entanto, a dificuldade em comparar sistemas para avaliar esse efeito tem sido um importante desafio. Neste estudo, foram produzidas e analisadas trajetórias de dinâmica molecular de duas variantes de ornitina transcarbamoilase (OTC), as quais apresentam arranjos estruturais semelhantes, mas apenas em uma é verificada a presença de um nó do tipo 3_1 . Os resultados da raiz do desvio quadrado médio (RMSD) mostraram estruturas equilibradas para as trajetórias produzidas e raiz da flutuação quadrática média (RMSF) mostrou diferenças sutis na flexibilidade. Na região do nó, a proteína com nó (PcN) não apresentou grande flutuação em nenhuma temperatura. Para a proteína sem nó, o resíduo GLY243 mostrou uma alta flutuação na região correspondente ao nó. A fração de contatos nativos (Q) apresentou perfil semelhante em todas as temperaturas, com maior diminuição para 436K. A investigação do comportamento conformacional por meio de análise de componentes principais (PCA) e mapa dinâmico de correlação cruzada (DCCM) mostrou que a PcN é menos propensa a sofrer alterações em sua conformação para as condições empregadas em comparação com a proteína sem nó. Os dados PCA mostraram que a proteína sem nó apresentou maior dispersão nas conformações, o que sugere que ela tem maior capacidade de transições de conformação em resposta às mudanças térmicas. Os gráficos de DCCM comparando as temperaturas 310K e 436K mostraram que a PcN teve menos mudança nos movimentos de correlação e anticorrelação, indicando maior estabilidade estrutural. Os descritores de reatividade locais usados na investigação de características da estrutura eletrônica mostraram diferenças para os resíduos idênticos nas duas proteínas na região do nó. Isso sugere que descritores de reatividade obtidos por cálculos quânticos podem ser usados em investigação do efeito que o nó pode exercer na estrutura eletrônica de PcN.

Palavras-chave: PcN. Dinâmica Molecular. PCA. DCCM. Descritores de Reatividade

ABSTRACT

The role of knots in proteins remains undefined. Some studies suggest an impact on stability, however, they have difficulties in comparing systems to assess this effect. In this study, were produced and analyzed molecular dynamic trajectories of two variants of ornithine transcarbamoylase (OTC), only one of which has a 3_1 knot, in order to evaluate the relative stability of the two molecules. Root mean square deviation (RMSD) showed equilibrated structures for the produced trajectories, and root mean square fluctuation (RMSF) showed subtle differences in flexibility. In the knot moiety, the knotted protein did not show a great deal of fluctuation at any temperature. For the unknotted protein, the residue GLY243 showed a high fluctuation in the corresponding moiety. The fraction of native contacts (Q) showed a similar profile at all temperatures, with the greatest decrease by 436K. The investigation of conformational behavior with principal component analysis (PCA) and dynamic cross-correlation map (DCCM) showed that knotted protein is less likely to undergo changes in its conformation under the conditions employed compared to unknotted. PCA data showed that the unknotted protein had greater dispersion in its conformations, which suggests that it has a greater capacity for conformation transitions in response to thermal changes. DCCM graphs comparing the 310K and 436K temperatures showed that the knotted protein had less change in its correlation and anti-correlation movements, indicating stability structural compared to the unknotted. The local reactivity descriptors used investigation of electronic structure characteristics showed differences for the identical residues in the knot moiety. This suggests that the reactivity descriptors obtained by quantum calculations can be used in the investigation of the effect that the node can have on the electronic structure of knotted protein.

Key words: Knotted Protein. Unknotted. PCA. DCCM. Reactivity Descriptors

Sumário

Agradecimentos	i
Resumo	ii
Abstract	iii
Lista de ilustrações	v
1 Introdução	15
2 Objetivos	19
2.1 Objetivo Geral	19
2.2 Objetivos Específicos	19
3 Fundamentação Teórica	20
3.1 Proteínas com Nó (PcN)	20
3.2 Caracterização computacional de Proteínas com Nó	22
3.3 Fundamentos de Dinâmica Molecular (DM)	26
3.4 Descritores de Reatividade	29
3.5 Simulação e Modelagem Molecular de Proteínas com Nó	32
4 Metodologia	41
4.1 Obtenção das Estruturas	41
4.2 Identificação do Nó na PcN	44
4.3 Critérios de Simulação Molecular	45
4.4 Análises de Estabilidade	47
4.5 Avaliação Conformacional por PCA e DCCM	49
4.6 Cálculo de descritores de reatividade	50
5 Resultados e Discussão	51
5.1 Identificação do nó em 3KZN	51
5.2 Análises de Estabilidade Estrutural	52
5.3 Análises do Comportamento Conformacional por PCA e DCCM	55
5.4 Descritores de Reatividade	72
6 Conclusão	84
Referências	85
A Apêndice	91

Lista de ilustrações

Figura 1 – Representação hierárquica da classificação estrutural de proteínas. Da esquerda para a direita a representação de estruturas: primária, secundária, terciária e quaternária	16
Figura 2 – Representação de uma <i>knotted protein</i> . Proteína com nó de entrada PDB 4JKA. Em verde a região do nó na estrutura, em vermelho os resíduos que delimitam o nó.	17
Figura 3 – Representação da estrutura da anidrase carbônica PDB 2CAB. Em violeta a região da estrutura que compreende o nó.	21
Figura 4 – Diagrama de representação da topologia de nós já encontrados nas proteínas.	22
Figura 5 – Representação de proteínas com nó (PcN). No topo a estrutura tridimensional, ao centro a representação obtida com uma suavização no servidor Knots para detecção do nó e na parte inferior a identificação do tipo de nó: 3_1 (PDB 1J85), 4_1 (PDB 1YVE), 5_2 (PDB 1XD3).	23
Figura 6 – Ilustração do procedimento de trabalho do algoritmo KMT para redução da configuração da estrutura da proteína.	24
Figura 7 – Métodos de fechamento utilizados para cadeias abertas de proteínas: a esquerda o fechamento aleatório usado no KnotProt e a direita um método alternativo, onde o fechamento do circuito ocorre pela menor distância entre os terminais.	25
Figura 8 – Ilustração das hipóteses de formação do nó por meio de estruturas obtidas por dinâmica molecular.	34
Figura 9 – A Estruturas representativas para a formação do nó 3_1 na proteína PDB 2EFV. B Evolução temporal obtida de trajetória DM com análise de formação do nó.	35
Figura 10 – Simulação da proteína MJ0366. A Número de contatos nativos x tempo de simulação por DM. B Mecanismo de desenvolvimento durante uma trajetória DM para uma $T=530K$. A perda do nó ocorre em 16.09 ns.	37

Figura 11 – Estrutura das proteínas: PDB 3KZN (PcN) em A e PDB 4JQO B	42
Figura 12 – Sobreposição das estruturas tridimensionais: A sobreposição de PDB 3KZN (PcN) em azul e PDB 4JQO em laranja; B destaque da região do nó e a região correspondente para a estrutura sem nó. Em C o alinhamento em relação a estrutura primária: em azul os resíduos conservados nas duas estruturas representam a baixa identidade (25%).	43
Figura 13 – Resultados obtido com o <i>KnotProt</i> para uma proteína com nó. O diagrama de refere a entrada PDB 5VIK.	45
Figura 14 – PcN 3KZN: A Destaque para a região do nó em magenta. B Resíduos da região do nó usados para avaliação dos descritores.	50
Figura 15 – Identificação do nó do tipo 3_1 na estrutura da proteína 3KZN. Na esquerda o diagrama representativo obtido com o <i>KnotProt</i> , a direita a representação tridimensional da estrutura com as cores correspondentes a informação do diagrama.	51
Figura 16 – RMSD e RMSF (50ns) para as proteínas PDB 3KZN (PcN) e 4JQO. Valores em azul e verde para 310K, vermelho e preto para 373K e roxo e amarelo para 436K. O retângulo demarca a região do nó na PcN. Os valores são referentes a triplicata.	53
Figura 17 – Fração de contatos nativos (Q) para 50ns de simulações de DM. A PcN é apresentada nas cores azul, vermelha e roxa para 310K, 373K, 436K, respectivamente. A proteína sem nó está representada em verde, preto e laranja para essas temperaturas. Valores referentes a triplicata	55
Figura 18 – PCA para a PcN (PDB 3KZN) em 310K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	57
Figura 19 – PCA para a PcN (PDB 3KZN) em 373K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	59

Figura 20 – PCA para a PcN (PDB 3KZN) em 436K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	61
Figura 21 – PCA para a proteína sem nó (PDB 4JQO) em 310K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	63
Figura 22 – PCA para a proteína sem nó (PDB 4JQO) em 373K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	65
Figura 23 – PCA para a proteína sem nó (PDB 4JQO) em 436K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em B as contribuições dos resíduos para as duas primeiras PCs.	67
Figura 24 – DCCM para a proteína sem nó (PDB 4JQO). Os dados são apresentados por temperatura de estudo: A 310K, B 373K e C 436K. Os retângulos destacam as áreas de mudança significativa.	69
Figura 25 – DCCM para a proteína com nó (PcN) (PDB 3KZN). Os dados são apresentados por temperatura de estudo: A 310K, B 373K e C 436K. Os retângulos destacam as áreas de mudança significativa.	70
Figura 26 – Distribuição de densidade de probabilidade para os descritores de reatividade calculados entre as estruturas amostradas para 3KZN (PcN) e 4JQO (sem nó). A) Nucleofilicidade, B) Eletrofilicidade, C) Netfilicidade, D) Moleza Local, E) Suscetibilidade radical, F) Dureza Local (aproximação potencial químico local), G) Localização do Orbital Molecular, H) Multifilicidade, I) Dureza Local (aproximação potencial de Fukui), J) Densidade Eletrônica, K) Dureza Local (aproximação interação elétron-elétron).	73

Figura 27 – Resultados da análise de componentes principais (PCA) empregando os dados obtidos pelos descritores de reatividade. A) Dados apresentados em relação ao tipo de proteína; B) Dados em relação ao tipo de resíduo de aminoácido; C) Variância explicada por componente principal; D) <i>Loadings</i> para os cinco primeiros componentes principais.	75
Figura 28 – Resultados da análise de componentes principais (PCA) para os descritores de reatividade. A) Objetos pelo tipo de proteína. B) Objetos pelo tipo de resíduo de aminoácido do nó.	76
Figura 29 – Mapas de calor mostrando as médias dos descritores para cada resíduo analisado A) PcN B) Proteína sem nó.	78
Figura 30 – Box-plot de alguns descritores químicos quânticos por resíduos de aminoácidos conservados no nó comparando entre os tipos de proteínas A) Moleza Local B) Dureza Local C) Localização de orbitais moleculares de fronteira D) Densidade eletrônica.	79
Figura 31 – Dureza local por átomos na região com nó em uma estrutura representativa: A) Dureza local com aproximação do potencial químico local para 3KZN (com nó) B) Dureza local com aproximação do potencial químico local para 4JQO (sem nó) C) Dureza local com aproximação interação elétron-elétron 3KZN (com nó) D) Dureza local com aproximação interação elétron-elétron 4JQO (sem nó). . .	80
Figura 32 – Dureza local por átomos, obtida a partir do potencial de Fukui. Resíduos em destaque para a região do nó em A) 3KZN (PcN) e B) para 4JQO. Em C) 3KZN e D) 4JQO são mostradas as representações considerando toda a estrutura das proteínas.	82

Introdução

Proteínas são macromoléculas largamente estudadas por cientistas experimentais e teóricos ao redor do mundo. Particularmente, estruturas com arranjo tridimensional complexo são um desafio na busca constante em compreender como a estrutura proteica impacta no funcionamento biológico^[1-4].

Essas estruturas desempenham papéis cruciais nos mais variados processos biológicos, entre os quais podem ser citados a extensa participação na catálise de reações químicas que ocorrem nos organismos, o transporte de moléculas como o oxigênio, indispensável ao funcionamento dos sistemas biológicos por meio da hemoglobina, mecanismos de defesas, os quais incluem imunoglobinas, e ainda, são essas estruturas moleculares que controlam a atividade dos genes por meio de proteínas reguladoras que se ligam ao DNA em sítios específicos^[5].

Em relação a composição molecular, as proteínas são polímeros de aminoácidos. Embora apresentem estruturas e funções tão diversificadas, a sequência de aminoácidos que compõem a estrutura é obtida combinando-se 20 tipos diferentes destes, os quais possuem em sua estrutura quatro grupos ligados a um átomo de carbono (carbono alfa, $C\alpha$), sendo eles um grupo amino ($-NH_2$), um grupo carboxílico ($-COOH$), um átomo de hidrogênio (H) e um grupo R, também denominado cadeia lateral, que é variável para cada aminoácido.

A compreensão da estrutura das proteínas se dá em diferentes níveis de complexidade. Essas macromoléculas são descritas em uma espécie de hierarquia conceitual, que normalmente é definida em quatro níveis: **estruturas primária, secundária, terciária e quaternária**^[6].

A figura 1 resume os diferentes níveis de classificação estrutural das proteínas.

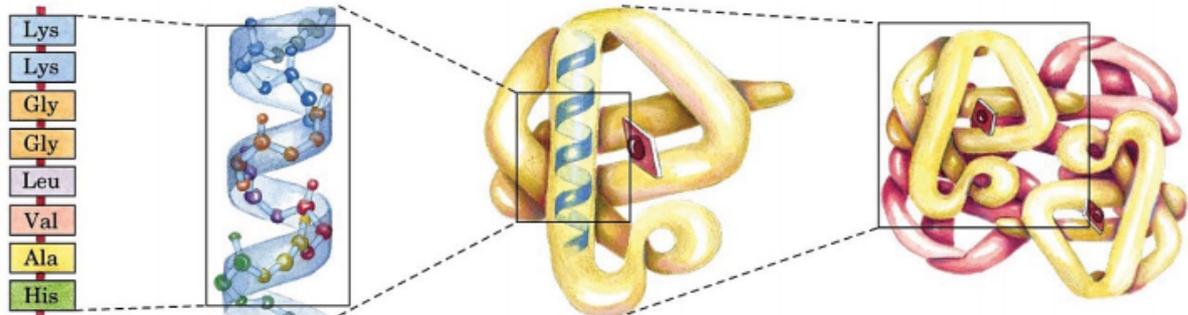


Figura 1 – Representação hierárquica da classificação estrutural de proteínas. Da esquerda para a direita a representação de estruturas: **primária, secundária, terciária e quaternária.**

Fonte: Adaptado de Francisco Junior e Francisco, 2006^[7].

Para que uma proteína exerça seu papel de forma adequada é necessário que esteja em sua forma biológica funcional específica, a qual recebe o nome de estado nativo. Esse por sua vez, coincide com o arranjo tridimensional da estrutura terciária da proteína (ou quaternária quando a proteína possui ao menos duas cadeias polipeptídicas associadas), sendo resultado do empacotamento das estruturas secundárias, como cadeias hélice α , folhas β , *loops*^[8,9].

Ao longo do tempo, estudos foram realizados com o intuito de desvendar a estrutura das proteínas e sua reatividade que é extremamente influenciada por seu arranjo espacial^[10].

Atualmente, embora já se saiba muito sobre a estrutura das proteínas, bem como as forças que as mantêm e sua atividade nos sistemas biológicos, há ainda proteínas que possuem uma característica marcante em seu arranjo estrutural, o que tem desafiado os pesquisadores quanto à sua função e surgimento – a presença de nós geométricos em sua estrutura^[4,11]. Devido a essa característica estrutural, essas macromoléculas são denominadas pelo termo em inglês *Knotted Proteins*.

A Figura 2 apresenta a estrutura tridimensional da proteína de entrada PDB 4JKA que tem um nó geométrico na sua estrutura.

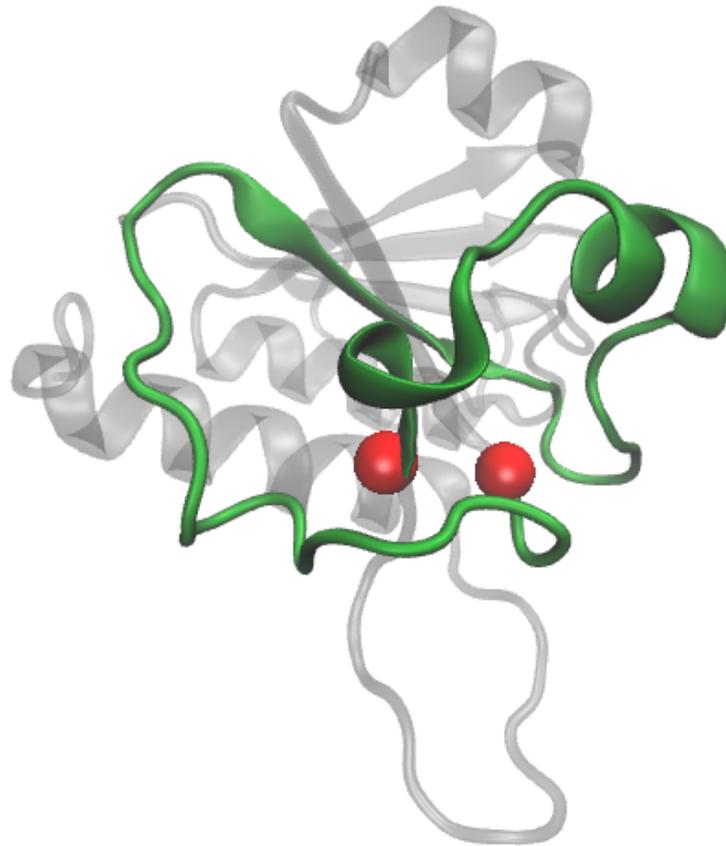


Figura 2 – Representação de uma *knotted protein*. Proteína com nó de entrada PDB 4JKA. Em verde a região do nó na estrutura, em vermelho os resíduos que delimitam o nó.

Fonte: criado com VMD^[12].

Nesse texto, essas estruturas serão tratadas como “Proteínas com Nó (PcN)”. Os esforços para compreensão do impacto que a presença de nó traz à proteína, bem como a busca por compreender como esses são formados têm se concentrado nos âmbitos experimental e teórico^[3,13–15].

O nó nas proteínas é identificado no estado nativo, ou seja, em sua conformação de atuação biológica^[4,14]. No entanto, a compreensão do papel deste arranjo para as propriedades estruturais ou de atuação nos processos biológicos permanece como um desafio para os pesquisadores^[16,17].

Embora o primeiro relato de nó em proteínas já tenha mais de 40 anos^[18,19], ainda hoje esse pode ser considerado um dos temas mais intrigantes em bioquímica estrutural^[17,20]. Com isso, estudos que possam contribuir para a compreensão desse tema, tendo em vista sua complexidade, podem abrir portas para investigação de impactos na relação estrutura-atividade de proteínas.

Neste trabalho buscou-se investigar características de duas variantes de ornitina transcarbamoilase (OTC) por meio de métodos computacionais, com foco em diferenças observadas para a proteína com nó.

Entre as proteínas com nós conhecidas, as OTCs tem o exemplo único que permite uma investigação da contribuição do nó para a estrutura, pois duas de suas variantes possuem arranjo tridimensional muito similar, mas em apenas uma se verifica a presença de um nó^[1,14] .

Capítulo 2

Objetivos

2.1 Objetivo Geral

Estudar proteínas com nós geométricos por meio de métodos de química computacional.

2.2 Objetivos Específicos

- Utilizar servidores online para identificação de proteínas com nó em sua estrutura;
- Aplicar Dinâmica Molecular para investigar propriedades dinâmicas de proteínas com nó;
- Realizar análises convencionais e outras com foco no comportamento dinâmico nas trajetórias produzidas por dinâmica molecular em proteínas com nós geométricos;
- Realizar uma análise comparativa entre variantes de transcarbamilase de ornitina (OTC) quanto a diferenças provenientes da presença de nó.
- Calcular descritores de reatividade para avaliar características eletrônicas de proteínas com nó.

Capítulo 3

Fundamentação Teórica

Neste capítulo abordamos conceitos que fundamentam a compreensão do tema estudado, apresentamos trabalhos já desenvolvidos sobre Proteínas com Nó, bem como os fundamentos dos cálculos usados na simulação das estruturas tratadas nesta Tese.

3.1 Proteínas com Nó (PcN)

A existência de nó na estrutura de proteínas durante muito tempo foi julgada como não sendo possível, tendo sido vista com certo ceticismo pela comunidade científica^[4,19].

A primeira afirmação da presença de um nó em uma proteína foi feita por Richardson em 1977^[18]. Em 1994 Mansfield^[19] realizou o primeiro levantamento de estruturas com nó no banco de dados de proteínas *Protein Data Bank*^{1[21]}. Nesse estudo, para um total de 400 estruturas analisadas, a presença de um nó foi observada somente em uma proteína, a anidrase carbônica (Figura 3).

¹[<https://www.rcsb.org/>](https://www.rcsb.org/)

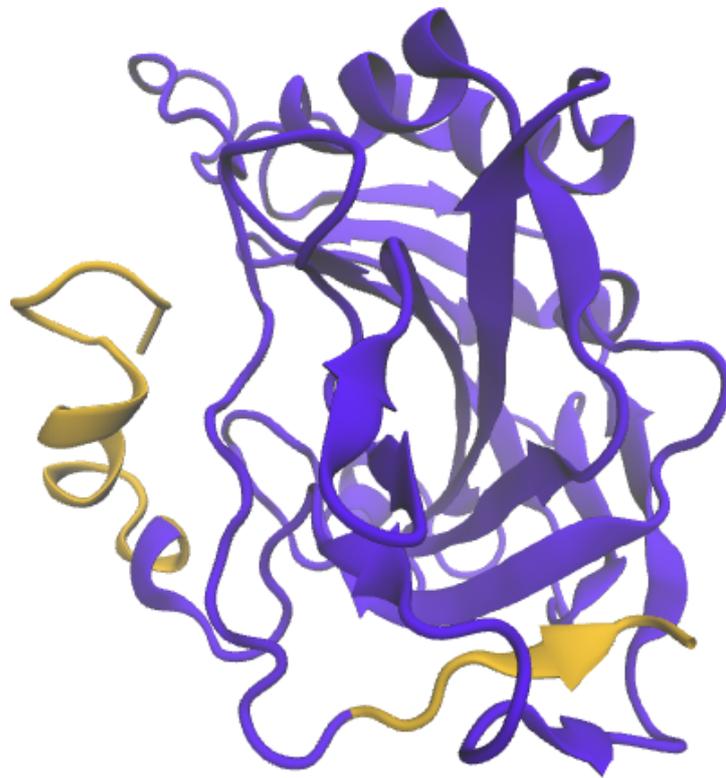


Figura 3 – Representação da estrutura da anidrase carbônica PDB 2CAB. Em violeta a região da estrutura que compreende o nó.

Fonte: criado com VMD^[12] .

Mesmo com os registros citados acima, de acordo com Jackson e colaboradores^[22] , até o ano de 2000 prevalecia o entendimento que as proteínas não deveriam apresentar nós em sua estrutura para que isso não comprometesse sua cinética durante o processo de enovelamento. Os nós seriam então evitados durante esse processo devido à cinética de sua formação ser complexa e lenta^[11,23] .

Nos últimos anos, estudos tem apontado que entre 1% e 2% das estruturas de proteínas depositadas no *Protein Data Bank* possuem pelo menos um nó^[11,13,24] . Considerando a enorme quantidade de estruturas já depositadas em banco de dados, essa pequena proporção de proteínas que possuem esse tipo de arranjo estrutural contribui para aumentar o interesse na investigação acerca do tema. O primeiro passo que permitiu a compreensão desse arranjo é a definição do que é um nó em proteínas, e como esse é determinado .

3.2 Caracterização computacional de Proteínas com Nó

A definição de nó é fundamentada através de um campo da matemática denominado de “Teoria dos Nós”^[25]. Nesta teoria, se estuda o emaranhamento em cadeias fechadas. Todavia, seus conceitos podem ser estendidos para estudar nós em cadeias abertas^[26].

O nó é definido como uma curva fechada no espaço que não possui autointersecção^[27]. Para essas curvas o nó é determinado de maneira única por meio de um invariante polinomial que é calculado sobre uma projeção da curva num plano. Entre os polinômios bem conhecidos dentro da teoria dos nós estão os polinômios de Jones e Alexander^[27-29], sendo este último o mais implementado nas ferramentas disponíveis para identificação de nós em proteínas^[16,26,30].

Os nós já caracterizados em proteínas são classificados de acordo com os cruzamentos das cadeias de polipeptídeos, que especificam sua topologia^[31]. O nó mais simples e que está presente na maior parte das proteínas é do tipo 3_1 , o qual possui três cruzamentos na cadeia. Em sequência de complexidade topológica tem-se os nós: 4_1 , que possui quatro cruzamentos, 5_2 que possui cinco cruzamentos de cadeias, e o nó do tipo 6_1 o de maior complexidade, possuindo seis cruzamentos^[9].

Na figura 4, abaixo são apresentados os diagramas representativos desses tipos de nós.

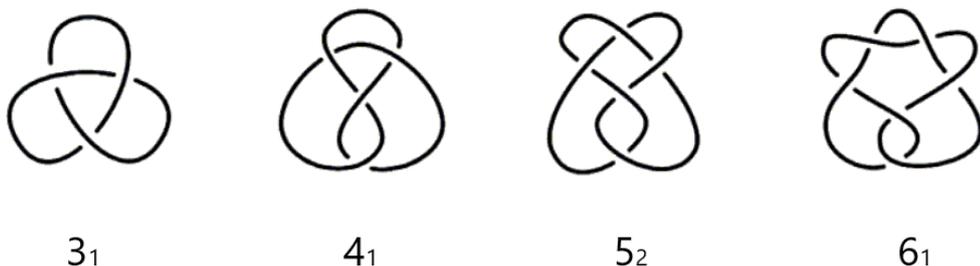


Figura 4 – Diagrama de representação da topologia de nós já encontrados nas proteínas.

Fonte: Adaptado de Alexander, Taylor e Dennis, 2017^[31].

Para a maioria das proteínas é muito difícil ou mesmo impossível visualizar a olho nu a presença de um nó em sua estrutura^[16,26]. Diante disso, diversos servidores e/ou programas fazem análise topológica dos nós nas estruturas a partir do arquivo

PDB ²[21] identificando o tipo e produzindo imagens do nó e de sua representação geométrica, além de fornecerem outras informações. Entre os servidores disponíveis, podem ser citados: KnotProt ³[16] e Knots ⁴[30]. Na figura 5 proteínas com diferentes nós são representadas.

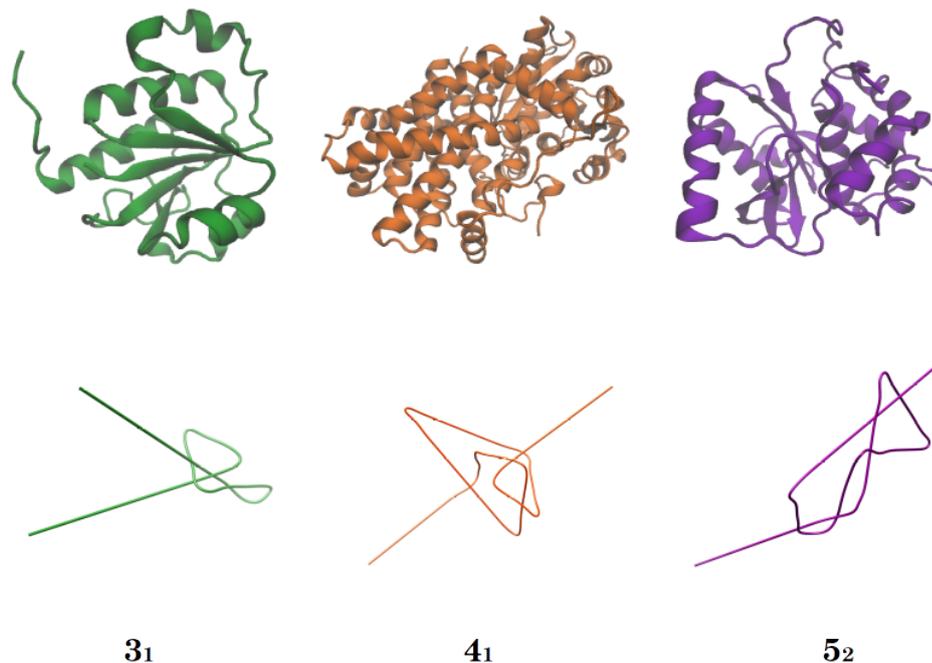


Figura 5 – Representação de proteínas com nó (PcN). No topo a estrutura tridimensional, ao centro a representação obtida com uma suavização no servidor Knots para detecção do nó e na parte inferior a identificação do tipo de nó: 3_1 (PDB 1J85), 4_1 (PDB 1YVE), 5_2 (PDB 1XD3).

Fonte: criado com VMD^[12] e UCFS Chimera^[32].

Conforme apresentado anteriormente, a Teoria dos Nós é aplicada para curvas fechadas. No entanto, em se tratando de proteínas que são formadas por cadeias abertas, para que essa teoria possa ser aplicada são utilizados argumentos matemáticos que normalmente permitem aplicar a teoria às cadeias polipeptídicas unindo-se seus resíduos terminais, tratando a cadeia principal como um circuito fechado^[33,34].

No procedimento para determinação dos nós nas proteínas, os servidores mais comuns empregam o algoritmo KMT^[29], que permite a redução da estrutura, mantendo somente os aminoácidos essenciais para a estrutura topológica da proteína^[26,30].

²[<https://www.rcsb.org/>](https://www.rcsb.org/)

³[<https://knotprot.cent.uw.edu.pl/>](https://knotprot.cent.uw.edu.pl/)

⁴[<http://knots.mit.edu/>](http://knots.mit.edu/)

O KMT^[29] trabalha de forma a avaliar todos os triângulos identificáveis numa cadeia formada por três aminoácidos em sequência, removendo o aminoácido médio, caso um triângulo não seja cruzado por nenhum outro segmento da cadeia. Depois de um certo número de iterações a cadeia é reduzida mantendo a mesma topologia da cadeia original. Esse processo é ilustrado na Figura 6.

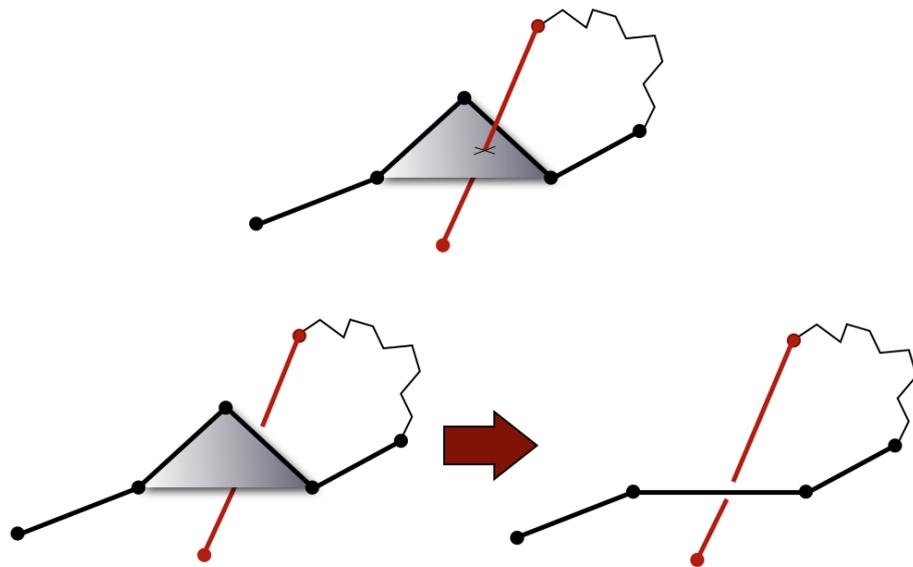


Figura 6 – Ilustração do procedimento de trabalho do algoritmo KMT para redução da configuração da estrutura da proteína.

Fonte: Adaptado da página do KnotProt^[16].

Especificamente no servidor *KnotProt*, com última versão *KnotProt 2.0* desenvolvida por Dabrowski-Tumanski e colaboradores^[16], após a redução da estrutura utilizando o algoritmo KMT, para o fechamento da cadeia é aplicado um método que conecta os pontos finais da estrutura centenas de vezes a dois pontos escolhidos aleatoriamente a partir de vértices de um icosaedro truncado. Estes dois pontos são então conectados por um arco na superfície da esfera, o tipo de nós que aparece com maior frequência entre as tentativas para uma determinada estrutura é então definido como seu tipo de nó. A figura 6, ilustra esse processo e também o método de fechamento determinístico.

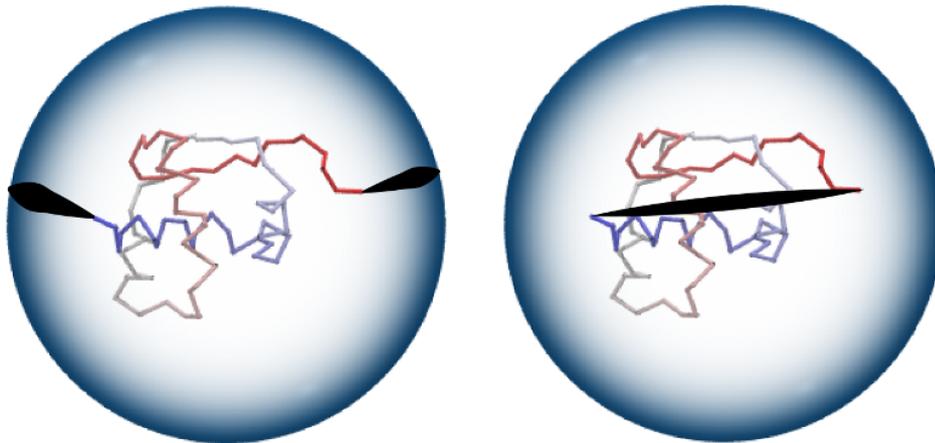


Figura 7 – Métodos de fechamento utilizados para cadeias abertas de proteínas: a esquerda o fechamento aleatório usado no KnotProt e a direita um método alternativo, onde o fechamento do circuito ocorre pela menor distância entre os terminais.

Fonte: Adaptado da página do KnotProt^[16].

O *KnotProt* apresenta os dados referente a topologia da proteína como um diagrama de matriz, definido como “*Impressão Digital*” para a PcN, mostrando detalhes como o tipo do nó, a composição deste, e ainda a presença de topologia “*slipknot*”. Devido a esta informação detalhada este é o servidor mais empregado para caracterização das proteínas com nó.

Apesar de a busca por desvendar as implicações do nó para a proteína ter tido avanços, ainda não é possível afirmar se estes exercem de fato alguma influência quanto ao desempenho da função biológica e se ocorrem acidentalmente, isto é, seu papel ainda não é claramente compreendido^[26,35,36].

Importante ao estudo das PcN é identificar entre estas qual o tipo de nó mais comum e a classe funcional que contém majoritariamente um tipo específico de nó. Pensando nisso, apesar do percentual baixo de PcN já identificadas frente ao total de estruturas de proteínas depositadas no *Protein Data Bank*^[21], alguns trabalhos apresentam esses dados que mostram os nós por meio das classes funcionais das proteínas^[22,37].

A partir dos dados já reunidos, usando como critério de busca, mostrar somente cadeias completas (todos os átomos) no servidor *Knotprot*^[16] a maior parte das PcN possui nó do tipo 3_1 . Para esse tipo são encontradas 484 entradas. Para os demais tipo de nó, o tipo 4_1 tem 37 entradas, para 5_2 são encontrados 10 entradas, e, para o tipo 6_1 3 entradas. Esses números são variáveis a depender do acesso, tendo em vista que o

servidor é atualizada semanalmente com base nas estruturas do *Protein Data bank*^[21] .

Pelo total de entradas no *Knotprot*,^[16] a maior parte das PcN possui nó do tipo 3₁, e para o tipo mais complexo (6₁), há somente uma proteína identificada, a α -haloácido dehalogenase (PBD 3BJX) . Essa proteína é um membro de uma importante classe de enzimas microbianas que catalisam a degradação de poluentes orgânicos por meio da clivagem da ligação carbono-halogênio^[38] .

Como já mencionado, o estudo de nós em proteínas, apesar de ter tido seu primeiro relato a mais de 40 anos, ainda requer avanços para que as implicações desse arranjo possam ser desvendadas. Nessa busca os estudos computacionais têm tido papel cada vez mais importante ao longo do tempo.

Trabalhos envolvendo técnicas de simulação e modelagem molecular de proteínas com nó tem buscado avaliar vários aspectos do ponto de vista teórico de como esses nós contribuem para o estado funcional das proteínas nos processos biológicos ou mesmo em relação a aspectos puramente estruturais^[1,11,39] .

Por essa razão, nas próximas seções serão apresentados os fundamentos da técnica de simulação molecular por Dinâmica Molecular (DM) que é a mais utilizada nos trabalhos que apresentam estudos sobre PcN. Além disso, foram apresentados os principais conceitos e uma descrição dos descritores de reatividade, que foram utilizados com o propósito de avaliar a estrutura eletrônica da proteína com nó e seu análogo sem nó, que compõem os sistemas estudados nesta Tese. E, em seguida, são apresentados trabalhos que trouxeram contribuições ao tema.

3.3 Fundamentos de Dinâmica Molecular (DM)

Por se tratar de um método que tem base na Mecânica Clássica, a dinâmica molecular utiliza as equações de movimento, primeiro propostas por Newton e depois generalizadas por Hamilton e Lagrange.

Pela Segunda Lei de Newton, o movimento da i -ésima partícula de um sistema, sujeita a uma força, é descrito conforme a Equação 3.1 :

$$F_i(t) = m_i a_i \quad (3.1)$$

Onde F_i é a força atuante sobre a i -ésima partícula do sistema, de massa m_i , em um

instante de tempo t , e a_i é a aceleração produzida por tal força na i -ésima partícula^[10]

Para se modelar moléculas grandes como as proteínas é imprescindível levar em consideração a interação entre os átomos que compõem o sistema, o que é possível por meio dos parâmetros que constituem o campo de força^[40].

Um campo de força é uma função de energia potencial que permite o cálculo da energia total do sistema através da sua estrutura tridimensional^[41]. O campo de força é obtido por meio da parametrização de equações matemáticas que reproduzem aspectos do comportamento molecular^[10], sendo constituído por termos referentes aos átomos ligados, tais como estiramento de ligação, deformação e torção angular, e termos representativos de interações entre partículas não ligadas, como interações de van der Walls e de Coulomb.

A Equação 3.2, representa um campo de força típico, onde o potencial total, $V(r)$, é resultado do somatório de potenciais para as interações existentes no sistema:

$$V(r) = \sum V_{(l)} + \sum V_{(\theta)} + \sum V_{(\varphi)} + \sum V_{(vdW)} + \sum V_{(elet)} \quad (3.2)$$

Nessa equação, V_l representa o estiramento de ligações, V_θ a deformação angular, V_φ o potencial de torção. Por fim, as interações referentes aos termos entre partículas não ligadas, interações de van der Walls e Coulomb, são computados em V_{vdW} e V_{elet} , respectivamente^[10].

Utilizando da aproximação por potenciais harmônicos, os termos referentes a estiramento e deformação angular são descritos pelas Equações 3.3 e 3.4:

$$V_{(l)} = \frac{1}{2}K_r(r - r_0)^2 \quad (3.3)$$

$$V_{(\theta)} = \frac{1}{2}K_\theta(\theta - \theta_0)^2 \quad (3.4)$$

Onde os parâmetros r e θ se referem ao comprimento e ângulo de ligação, tendo o valor r_0 e θ_0 em torno da posição de equilíbrio, respectivamente. Os termos representados por K_r e K_θ são as constantes de força para o sistema^[40].

A representação do potencial energético usado para descrever a torção, tem forma de acordo com a Equação 3.5^[41]:

$$V_{(\varphi)} = \frac{V_n}{2}(1 + \cos(n\varphi) - \gamma) \quad (3.5)$$

Onde V_n , n , φ e γ representam, respectivamente, a barreira energética imposta a torção, número de máximos ou mínimos para uma torção completa, o ângulo diedro e o ângulo de fase. Alguns campos de força trazem ainda aproximações harmônicas usadas para descrever torções impróprias^[41]. Estas torções são observadas para o arranjo de quatro átomos, onde um átomo central está ligado a outros três.

As contribuições referentes as interações entre partículas não ligadas, são inseridas por meio dos potenciais de Lennard-Jones (V_{vdW}) e Coulomb (V_{elet}), representados pelas Equações 3.6 e 3.7:

$$V_{(vdW)} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (3.6)$$

$$V_{(elet)} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (3.7)$$

Os termos que compõem esses potenciais são a profundidade (ϵ) e largura (ψ) do poço e a separação entre as partículas (átomos) representados por r_{ij} . Na equação para o potencial eletrônico (V_{elet}), os termos das cargas pontuais são representados por q_i e q_j , ϵ_0 é a permissividade no vácuo, e ϵ_r a constante dielétrica relativa do meio^[41].

Diversos campos de força estão disponíveis para a modelagem de proteínas, entre os mais usados podemos citar AMBER, CHARMM e GROMOS^[40]. Definindo-se o campo de força a ser utilizado, o próximo passo consiste na propagação das equações de movimento para o sistema em estudo.

A força que atua sobre cada átomo é determinada pela derivada do potencial para cada átomo após a escolha do campo de força a ser utilizado^[41]. Assim tem-se a equação diferencial 3.8 que gera diretamente a força atuando sobre cada partícula, a cada passo da evolução no tempo:

$$\vec{F}_i(t) = \left(\frac{\partial V_i(\vec{r}_i)}{\partial \vec{r}_i} \right) \quad (3.8)$$

Uma vez obtidas as forças, a equação proveniente da Segunda Lei de Newton pode ser integrada no tempo para a obtenção das novas velocidades e as mudanças nas posições de cada átomo. De posse, dessas novas posições e velocidades para cada

partícula do sistema determina-se as energias potencial e cinética. Esse procedimento é repetido sucessivas vezes para que se obtenha o conjunto de velocidades e posições de cada partícula ao longo do tempo^[41].

Existem vários algoritmos que são comumente empregados na integração das equações de movimento da mecânica clássica. Um dos mais importantes é o algoritmo de Verlet que é extensivamente utilizado em simulações de Dinâmica Molecular. A Equação 3.9 descreve o cálculo realizado pelo referido algoritmo

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \quad (3.9)$$

No cálculo da equação 3.17 são consideradas as posições e acelerações do tempo (t), e as posições do passo anterior, definido por $r(t - \Delta t)$ para determinar novas posições no tempo $t + \Delta t$.

Para se chegar a equação anterior é feita a variação das posições, variando a equação "para frente e para trás", por meio de uma expansão de Taylor, de acordo com as Equações 3.10 e 3.11, respectivamente^[41],

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \dots \quad (3.10)$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \dots \quad (3.11)$$

Somando-se essas duas equações, e isolando o termo $r(t + \Delta t)$, se obtém a equação do algoritmo de Verlet.

3.4 Descritores de Reatividade

Embora a maioria dos trabalhos encontrados para o estudo de proteínas com nó (PcN) empreguem dinâmica molecular, características provenientes da estrutura eletrônica não podem ser acessadas por meio dessa técnica que emprega potenciais clássicos. Para tanto, são necessários métodos que empreguem química quântica, permitindo tratar a estrutura eletrônica dos sistemas.

Um dos poucos exemplos encontrados na literatura que utilizou métodos de química quântica é o estudo recente de Perlinska e colaboradores^[42]. Nesse trabalho foi investigada atuação do Mg^{2+} na catálise envolvendo uma PcN. No entanto, não

foram exploradas as características ou contribuições do nó para a estrutura em relação a reatividade.

Biomoléculas são desafiadoras quanto o estudo de características da estrutura eletrônica devido a elevada quantidade de elétrons a serem tratados. Recentemente o uso de Descritores de Reatividade (DR) foi apresentado com sucesso para avaliar tais características^[43-45]. Esses descritores podem ser obtidos empregando hamiltonianos semiempíricos que potencializam o uso de recurso computacional^[46].

Descritores de Reatividade são quantidades teóricas utilizadas para descrever resumidamente propriedades de reatividade de sistemas moleculares com base em sua densidade eletrônica^[46,47]. São obtidos por meio de manipulações matemáticas de resultados provenientes de estruturas eletrônicas convergentes, como por exemplo, coeficientes de orbitais moleculares^[47]. Entre estes, os mais usados são definidos pela Teoria Funcional da Densidade Conceitual (do inglês *Conceptual Density Fuction Theory*, CDFT)^[46,48].

Dentro da CDFT, os descritores de reatividade são definidos pela equação diferencial total da DFT (Equação 3.12) e suas derivadas^[49], que relaciona a energia eletrônica (E) com o potencial químico μ , a densidade eletrônica ($\rho(r)$) para o estado de referência e o potencial externo ($v(r)$)

$$dE = \mu dN + \int \rho(r) dv(r) dr \quad (3.12)$$

Dentre as possibilidades de investigação empregando DRs estão a dureza e moleza química de Pearson, que definiu os processos de interação duro-duro como controlados por carga e mole-mole por controlados por orbital^[50], eletronegatividade, susceptibilidade de ataques nucleofílicos e eletrofílicos^[46], quantidades importantes para a investigação da reatividade de sistemas químicos.

Neste estudo, usamos o software PRIMoRDIA^[47], que permite calcular uma série de descritores de reatividade para sistemas de muitos átomos como as proteínas. Utilizando esse software alguns fenômenos já foram investigados com sucesso, tais como elucidção da catálise enzimática por meio do delineamento do perfil de reação com métodos de baixo custo computacional^[43], interação proteína-ligante para o sistema HIV1-protease^[43], inibidores da toxina da ricina A (RTA)^[44], impacto da variação da estrutura eletrônica na reatividade de M^{pro} dos vírus SARS-CoV-I e SARS-

CoV-II^[45] .

Neste trabalho os descritores de reatividade calculados, foram empregados para explorar características da estrutura eletrônica de uma PcN (AOTCase) em relação a seu análogo sem nó (OTCase). Entre dos descritores disponíveis no PRIMoRDiA^[47] , destacamos a seguir os que foram selecionados para explorar as características eletrônicas dos dois sistemas em estudo.

Para avaliar diferenças provenientes da estrutura eletrônica, foram usados descritores de reatividade locais. Tendo em vista, que neste estudo o interesse está diretamente atrelado a região do nó, com os descritores locais é possível avaliar diferenças entre as duas moléculas com observações locais de reatividade.

Entre os diversos descritores de reatividade disponíveis no software PRIMoRDiA foram selecionados a eletrofilicidade, nucleofilicidade, dureza e moleza locais, suscetibilidade radical, localização do orbital molecular, multifilicidade e densidade eletrônica.

Os descritores de reatividade locais são matematicamente definidos por uma função no espaço tridimensional, r , isto é, para um dado ponto no espaço o descritor assume certo valor^[47] . Por definição, a densidade eletrônica ρ é uma quantidade local, onde o seu valor é definido pelo produto dos orbitais moleculares ocupados $\phi_{occ}(r)$, que são combinações lineares de orbitais atômicos ϕ , onde os coeficientes c_i são determinados por métodos computacionais.

A densidade eletrônica é definida pela Equação 3.13. Os orbitais moleculares são determinados de acordo com a Equação 3.14.

$$\rho(r) = \sum_i^{occ} |\psi(r)_i|^2 \quad (3.13)$$

$$\psi(r)_i = \sum_i c_i \phi(r) \quad (3.14)$$

A partir da densidade eletrônica (que substitui o número de elétrons usado para o cálculo dos descritores globais) é que se realiza os cálculos dos descritores de reatividade locais que permitem avaliar a estrutura eletrônica para as macromoléculas, como as proteínas usadas nesse estudo. Dados como, por exemplo, a densidade eletrônica total, orbitais moleculares (MOs) e o potencial eletrostático molecular são obtidos automaticamente quando o PRIMoRDiA é executado^[47] .

Os descritores nomeados como eletrofilicidade e nucleofilicidade podem ser interpretados como capacidades de acomodar e doar carga, respectivamente, por um sistema em estudo^[43,47]. Assim, fornecem informações sobre o átomo ou resíduo acerca dessas características em relação a distribuição da sua densidade eletrônica. Dureza e moleza locais são grandezas químicas já conhecidas, que fornecem informações sobre o tipo de interação favorável as espécies químicas analisadas (interações duro-duro e mole-mole).

Usando o PRIMoRDIA calculamos os descritores de dureza e moleza locais por diferentes métodos de cálculo^[43,47], como potencial químico eletrônico, potencial de Fukuiquímico local (LCP) e potencial de interação elétron-elétron.

Outros descritores são determinados a partir dos valores da eletrofilicidade e a nucleofilicidade. A netfilicidade é o resultado da diferença entre esses dois descritores, enquanto que a suscetibilidade radical (ou radicalidade) é o valor obtido para a média desses. Além desses, há a multifilicidade que é determinada pela netfilicidade multiplicada pela eletrofilicidade total^[47].

3.5 Simulação e Modelagem Molecular de Proteínas com Nó

Muitos dos trabalhos que inicialmente investigaram proteínas com nó usaram modelos simplificados de estrutura, como *Coarse-Grained* (CG) para tentar compreender como o nó é formado^[36,38,51,52].

Nesse modelo de representação vários átomos próximos são agregados em uma única partícula^[10]. Isso tem impacto direto na demanda computacional facilitando a manipulação de sistemas que naturalmente contém muitos átomos, como é o caso das proteínas. Por outro lado, o uso de um modelo simplificado como o CG impede o detalhamento da estrutura que é importante para compreender a fundo o funcionamento da estrutura proteica da forma mais completa possível.

Em 2009, Sulkowska e colaboradores^[51] fizeram um estudo das proteínas YibK (PDB 1J85) e YbeA (PDB 1O6D) que contém nó do tipo 3_1 em sua estrutura, tendo como propósito avaliar qual a via de formação do nó no processo de enovelamento das estruturas. Nesse trabalho os autores usaram modelos simplificados de estrutura (CG) para realizar uma simulação por dinâmica molecular (DM), testando três diferentes

hipóteses de caminho para que as estruturas atingissem seu estado nativo.

Para as duas estruturas avaliadas os resultados apresentados apontaram que o nó é formado a partir de uma configuração inicial de *slipknot*, o que na verdade consiste numa conformação onde a proteína possui dois cruzamentos em suas cadeias, a partir desta ocorrendo o terceiro cruzamento no processo de enovelamento tem-se a formação do nó 3_1 .

Os passos são descritos por meio de seis configurações distintas obtidas com a DM: durante a primeira transição é criado um *loop*, do qual outra região da cadeia polipeptídica se aproxima logo em seguida, forçando uma conformação do tipo gancho. Esse gancho passa então através do *loop*, gerando dois cruzamentos das cadeias. Por fim, o terminal da proteína é puxado através do *loop* formando o nó do tipo 3_1 ^[51]. A limitação no entanto, surge pelo tipo de modelo usado, onde a estrutura proteica é reduzida a representação apenas da cadeia principal.

Muitos dos trabalhos observados na literatura buscam elucidar o mecanismo de formação de nós do tipo mais simples (3_1). No entanto, a proteína que possui o nó de topologia mais complexa identificada até o momento, a α -haloácido dehalogenase (PDB ID 3BJXj) com topologia do nó 6_1 , também já foi objeto de estudo por meio de simulações por DM empregando modelo de estrutura (CG)^[38]. Os autores apontaram as limitações que tal representação estrutural traz, mas destacaram a importância do trabalho devido à complexidade do nó presente na estrutura estudada.

O tamanho e complexidade da proteína foi relatado como fator que impossibilitou o estudo completo do processo de modo que pudesse inserir dados de natureza termodinâmica, o que permitiria construir o perfil de energia livre ao longo de uma coordenada de reação. Contudo, usando dos dados cinéticos distinguiram características para propor um provável caminho para a formação do nó no enovelamento. Foram computadas um total de 1000 trajetórias, dentre as quais seis trajetórias de sucesso foram obtidas (houve a formação do nó na estrutura ao término da DM), todas apresentando comportamento semelhante.

O resumo de uma trajetória estudada é representado pelas mudanças na estrutura da proteína, conforme ilustrado na figura 8, abaixo.

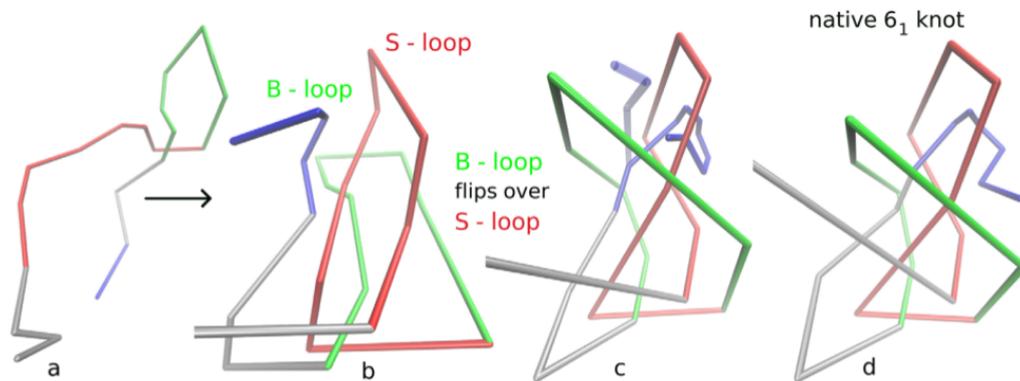


Figura 8 – Ilustração das hipóteses de formação do nó por meio de estruturas obtidas por dinâmica molecular.

Fonte: Adaptado de Boliölinger^[38].

Duas hipóteses foram testadas para a formação do nó nas estruturas avaliadas. Os passos iniciais são os mesmos para ambas, que consistem em formação de dois *loops* (representados na figura como *B-loop* e *S-loop*) por meio de torções na proteína parcialmente desdobrada. A partir disso as vias se distinguem.

Em uma via, ocorre a passagem do C-terminal (ilustrado em azul) pelo *S-loop* (com isso ocorre mudança na conformação desse *loop*), para em seguida o *B-loop* passar sobre o *S-loop* (via direta $0_1 \rightarrow 6_1$). Na segunda via, o *B-loop* realiza um giro sobre o *S-loop* e o C-terminal (via com intermediário $0_1 \rightarrow 4_1 \rightarrow 6_1$), sendo esta de acordo com o trabalho a via de enovelamento que explica a formação do nó. Para essa conclusão foram analisados os tempos para cada um dos fenômenos descritos, tendo sido encontrado maior correlação entre o tempo total do enovelamento versus os tempos de cada fenômeno para as trajetórias do caminho $0_1 \rightarrow 4_1 \rightarrow 6_1$.

Em trabalho publicado por Noel e colaboradores^[52] avaliou-se o sucesso de simulações em solvente explícito em comparação a modelos baseados em estrutura (SBM, do inglês *structure-based models*) para formação de um nó do tipo 3_1 para a proteína PDB 2EFV. Os pesquisadores utilizaram estruturas intermediárias obtidas por SBM selecionadas de forma aleatória contendo 10 resíduos encadeados do total de 15 que compõem a estrutura do nó da proteína para avaliar a formação do nó empregando DM com solvente explícito.

O cálculo realizado consistiu em uma dinâmica de equilíbrio simples utilizado o campo de força Amber99sb *all atom* (contabiliza todos os átomos da estrutura) e modelo de solvente TIP3P, tendo sido geradas 15 trajetórias, totalizando 25 μ s, usando

o software GROMACS v3.3.

No estudo, os autores avaliaram o terceiro cruzamento das cadeias polipeptídicas que confere a estrutura a condição de proteína com nó, como pode ser observado na figura 9A (etapa B-C, resíduo ARG91). Na parte B da figura 9 é apresentada a evolução temporal dos primeiros 350 ns da DM em relação aos parâmetros raio de giro (R_g), fração de contatos nativos (Q), SK_N e SK_C , referentes ao primeiro e último resíduo em $\alpha 4$ (figura 9A), participantes do cruzamento.

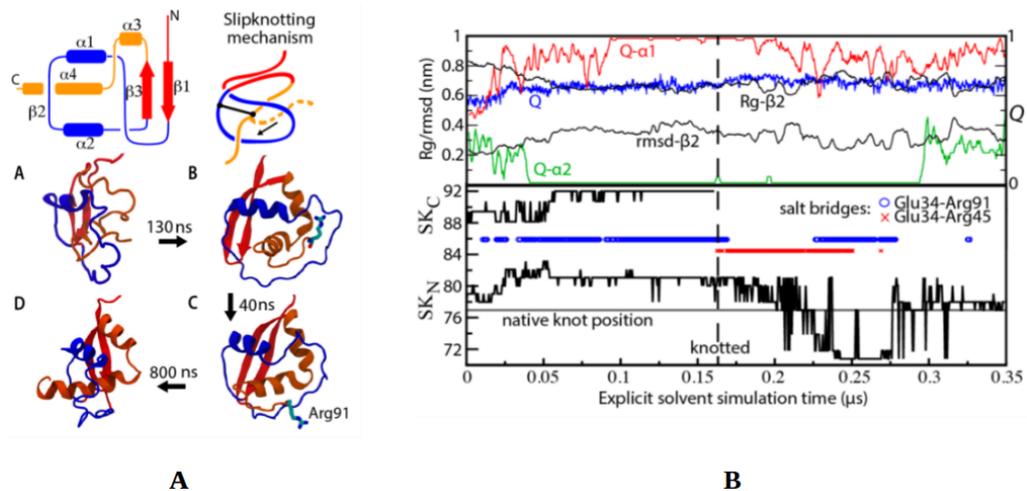


Figura 9 – A Estruturas representativas para a formação do nó 3_1 na proteína PDB 2EFV. B Evolução temporal obtida de trajetória DM com análise de formação do nó.

Fonte: Adaptado de Noel e colaboradores^[52].

Como se pode observar o terceiro cruzamento que caracteriza a formação do nó na estrutura ocorre quando o resíduo ARG91 cruza o *loop* (em azul) demarcado em 40 ns (transição das estruturas B→C).

Na parte B da figura 9 o progresso da dinâmica molecular realizada apresenta o comportamento da estrutura durante o processo de formação do nó: na parte superior a fração de contatos nativos para cada proteína esquematizada (ver parte superior da figura 9-A, trajetória em azul). Conforme essa fração aumenta (0,55 para 0,73), a estrutura atinge o estado nativo; $Q-\alpha 1$ e $Q-\alpha 2$, demonstram o comportamento das cadeias envolvidas na formação do nó durante o processo de enovelamento.

Quando SK_C se torna indefinido a proteína encontra-se com nó. A flutuação apresentada no comportamento de SK_N se deve a movimento de $\alpha 4$ que se encontra no

núcleo hidrofóbico, de modo que a falta de especificidade de interações hidrofóbicas poderia explicar a flutuação.

Embora os autores tenham usado um modelo de estrutura com todos os átomos da proteína, avaliaram que não houve diferença significativa do mecanismo já estudado usando modelos simplificados em trabalhos anteriores^[52].

Com o intuito de investigar o nó do tipo 3_1 no processo de enovelamento de proteínas, Wu e colaboradores^[53] usaram simulações de dinâmica molecular (DM), além de outras estratégias para investigar as proteínas MJ0366 (PDB 2EFV), presente em *Methanocaldococcus jannaschii*, e VirC2 (PDB 2RH3), presente no microorganismo *Agrobacterium tumefaciens*. As simulações de DM foram realizadas com o pacote GROMACS 4.0.7, utilizando campo de força GROMOS.

Para MJ0366 foram realizadas nove trajetórias em alta temperatura, definida em 530K, tendo em vista que em altas temperaturas ocorre o seu desenovelamento, e uma para o estado nativo, em 298K, para fins de comparação. O tempo total para cada simulação foi de 40 ns. A estrutura obtida pelo PDB foi solvatada com o modelo de água SPC216, após isso foi feita a minimização da estrutura e, em seguida, o sistema foi submetido a simulações de dinâmica molecular com controles da pressão e temperatura (detalhes na referência^[53]).

A proteína MJ0366 foi avaliada pela fração de contatos nativos, que consiste na razão entre o número de contatos nativos entre estrutura desenovelada e a estrutura nativa (enovelada). Na figura 10A, a fração de contatos nativos demonstra que a estabilidade da folha β remete a estabilidade da proteína como um todo. Dessa forma pelo seu desenovelamento, todo o sistema teria maior facilidade em desenovelar, e com isso, o nó presente na estrutura é desfeito. Na simulação por DM (trajetória apresentada na figura 10B) no tempo de 16,09 ns a proteína perde o nó e não volta a apresentá-lo em sua estrutura^[53].

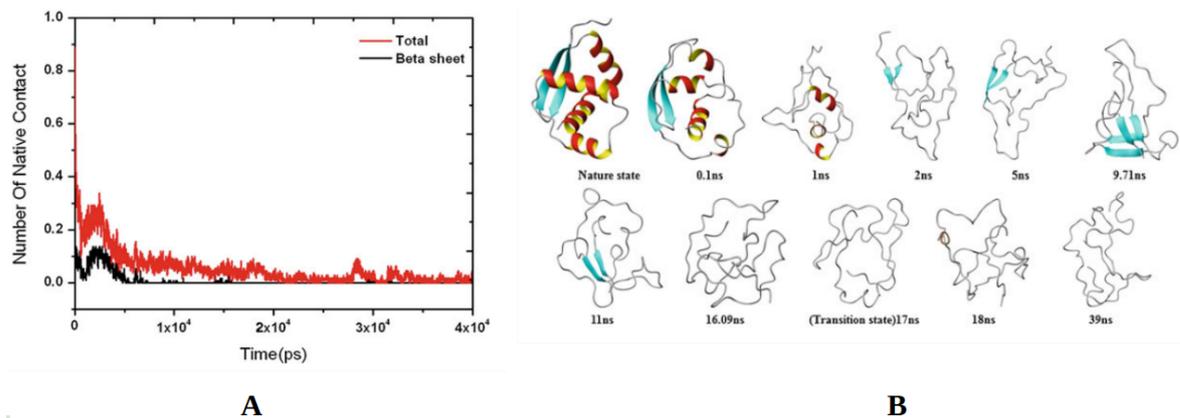


Figura 10 – Simulação da proteína MJ0366. **A** Número de contatos nativos x tempo de simulação por DM. **B** Mecanismo de desenvolvimento durante uma trajetória DM para uma $T=530\text{K}$. A perda do nó ocorre em 16.09 ns.

Fonte: Adaptado de Wu e colaboradores^[53].

Para a proteína VirC2, o estudo foi realizado com um modelo simplificado, onde os autores também encontraram relevância da folha β para a estrutura contendo nó.

Como mencionado anteriormente, embora a maior parte dos estudos computacionais sobre proteínas com nós se concentre em investigar proteínas que apresentam nó do tipo 3_1 , devido sua menor complexidade em relação aos demais, os demais tipos se apresentam como grandes desafios. A exemplo a única proteína que contém nó 6_1 citada anteriormente e o tipo 5_2 , a segundo mais complexo já identificado.

Em trabalho recente, Zhao e colaboradores^[36] estudaram estruturas que possuem nó do tipo 5_2 . Utilizando estruturas baseadas no modelo CG, foram realizadas simulações de dinâmica molecular usando o pacote GROMACS 4.5.4, tendo sido realizado um total de 200 simulações para cada estrutura utilizada no trabalho. As proteínas correspondem a família UCH, com entradas PDB 3IRT, 2LEN, 4I6N. Conforme justificam os autores, essas estruturas possuem caminho de enovelamento similares, por essa razão concentraram as análises na estrutura comumente estudada, UCH-L1 (PDB 3IRT).

Para o processo de enovelamento, acompanhado pela evolução temporal da estrutura, ocorre inicialmente o colapso desta, sendo observadas duas vias topológicas distintas para a formação do estado nativo que contém o nó 5_2 , diferindo na ordem de formação das estruturas secundária e terciária referentes aos terminais C e N: em uma das vias após o colapso da estrutura ocorre a transição direta do estado dese-

novelado para o estado nativo com nó ($0_1 \rightarrow 5_2$), onde o N-terminal é estruturado por último, após o C-terminal e, na outra via ocorre o contrário, com o C-terminal sendo estruturado por último, a proteína passa por um intermediário que contém um nó do tipo 3_1 ($0_1 \rightarrow 3_1 \rightarrow 5_2$), formação surge quando o N-terminal já se encontra estruturado e o C-terminal não^[36].

Por meio da determinação da probabilidade de ocorrência das vias, foi estimado que a formação do nó em via direta é predominante em relação a formação de um intermediário com topologia 3_1 . Isso decorre do fato do C-terminal formar a cadeia beta, sendo mais fortemente puxado em relação ao estado nativo, quando em comparação com o N-terminal solto. Em termos energéticos, a energia potencial média quando apenas o C-terminal está estruturado é sempre menor do que quando somente o N-terminal está estruturado.

Na tentativa de compreender as implicações que a presença de um nó pode trazer para a proteína, uma via de investigação para as proteínas com nós tem sido comparar alguns parâmetros destas com proteínas similares que não tenham nó em sua estrutura.

Em um estudo realizado por Xu e colaboradores^[3], os autores compararam a proteína YbeA proveniente do *Escherichia coli*, que possui um nó do tipo 3_1 , com uma proteína modificada YbeA* (a mesma proteína, tendo sido removido o nó da estrutura pela inversão de um cruzamento). O estudo foi realizado por meio de simulações de dinâmica molecular (DM) realizadas com o pacote GROMACS 4.6.7, empregando campo de força CHARMM e modelo TIP3P para a água. A simulação teve um tempo total de 100 ns.

Foram aplicadas aos dois sistemas estudados condições desnaturantes química (adicionando moléculas de uréia), térmica (impondo a condição de temperatura elevada, 520 K) e mecânica (aplicando uma força externa para o alongamento do C-terminal mantendo o N-terminal fixo). A resposta dos sistemas frente as condições desnaturantes foi analisada durante a evolução temporal da dinâmica molecular realizada^[3].

Os resultados obtidos mostraram que, sob o agente desnaturante químico, a estrutura que não continha o nó perdeu a estrutura secundária com 50 ns de simulação, enquanto que a estrutura com nó no mesmo tempo teve a estrutura secundária pouco

perturbada. Em relação ao agente térmico, após a elevação da temperatura, foi avaliada a fração de contatos nativos para verificar a desnaturação das estruturas durante a simulação. A proteína YbeA* foi completamente desnaturada com apenas 20 ns, enquanto que a mudança para YbeA foi muito mais lenta e de modo gradual, indicando que várias barreiras existem no processo para perda da estrutura enovelada da proteína que possui nó^[3].

Por fim, avaliando a desnaturação devido a ação de agente mecânico, onde uma força de 400 kJ/mol/nm foi empregada para realizar um estiramento na estrutura, observou-se para a estrutura com nó maior resistência ao estiramento (a distância final de ponta a ponta foi de 20 nm para YbeA e 16 nm para YbeA*), o que aponta para maior estabilidade dessa proteína, inferindo-se que o nó propicia resistência adicional frente a ação do agente mecânico de desnaturação^[3].

A busca por vantagens funcionais da presença do nó em proteínas foi investigada para proteínas que possuem em sua estrutura nó do tipo 3_1 . Foram analisadas as proteínas TrmL (PDB 4JAK), AOTCase (PDB 3KZN) e Rds3p (PDB 2K0A)^[1]. Nesse estudo foi utilizada ainda a proteína OTCCase (PDB 4JQO), que não possui nó em sua estrutura, para um comparativo com a AOTCase, tendo em vista que ambas pertencem a classe das tRNA-metiltransferases e possuem estrutura e função muito semelhantes). Foram utilizados na investigação mapas de contatos obtidos por meio do servidor Frustratometer, determinação do fator-B para análise da estabilidade estrutural e o índice SASA para a acessibilidade do solvente, ambos obtidos utilizando o software Chimera.

Para as três proteínas com nós verificou-se propriedades similares que os autores correlacionaram a presença de nós nas estruturas, ressaltando que estas não são relacionadas em sequência e função. Para todas foi constatado aumento acentuado do número de contato nas bordas do núcleo do nó. Esse número de contatos intracadeia elevado por sua vez conduz a maior estabilidade térmica e diminuição na acessibilidade do solvente a estas regiões.

Na comparação entre AOTCase e OTCCase, os resultados indicaram regiões muito semelhantes em relação ao aumento do número de contatos. Porém, os autores salientam que a estrutura que não possui nó (OTCase) mantém em seu arranjo dois dos três cruzamentos que caracterizam o nó, o que provavelmente dificultou a distinção

pretendida.^[1]

Alguns trabalhos trazem ainda a possibilidade de proteínas com nó terem também maior resistência quando a degradação controlada de proteínas não funcionais, o que se apresenta como ponto importante já que este é um dos processos cruciais nas células^[54-56] .

Recentemente, em um estudo com diferentes topologias de nós em proteínas, foi observada maior resistência à degradação em comparação com variantes sem nós.^[57] Diferentes tipos de nós têm efeitos diferentes em relação à degradação, que pode ocorrer de forma parcial ou com formação de intermediários^[58] . Nesta etapa, o nó tem influência quando em contato com o canal responsável pela degradação, como observado na protease AAA+ ClpXP^[55] , e um efeito de mecanoestabilidade já foi observado para proteínas com nó^[54,59] . Isso reafirma o interesse em estabelecer o impacto que a presença do nó traz para as proteínas.

Metodologia

4.1 Obtenção das Estruturas

As estruturas utilizadas para este trabalho foram obtidas no banco de dados de proteínas *Protein Data Bank*^[21]. As duas estruturas estão identificadas com os códigos PDB 3KZN e 4JQO, sendo estas um valioso exemplo de proteínas pertencentes a mesma família, com arranjo tridimensional similar, mas que em apenas uma se verifica a presença de nó^[14]. Essas entradas se referem as proteínas N-acetilornitina transcarbamoilase (ATOCase) e ornitina transcarbamoilase (OTCase), as quais foram utilizadas por Dabrowski-Tumanski e colaboradores^[1] em um trabalho que buscou encontrar diferenças entre as suas estruturas que fossem provenientes da presença de nó.

A figura 11 apresenta as estruturas das proteínas utilizadas neste estudo. Na parte **A** (AOTCase, PDB 3KZN), a proteína com nó e na parte **B** (OTCase, PDB 4JQO) a estrutura análoga que não contém nó em seu arranjo tridimensional. Cada proteína possui duas regiões de folha β espaçadas de modo similar.

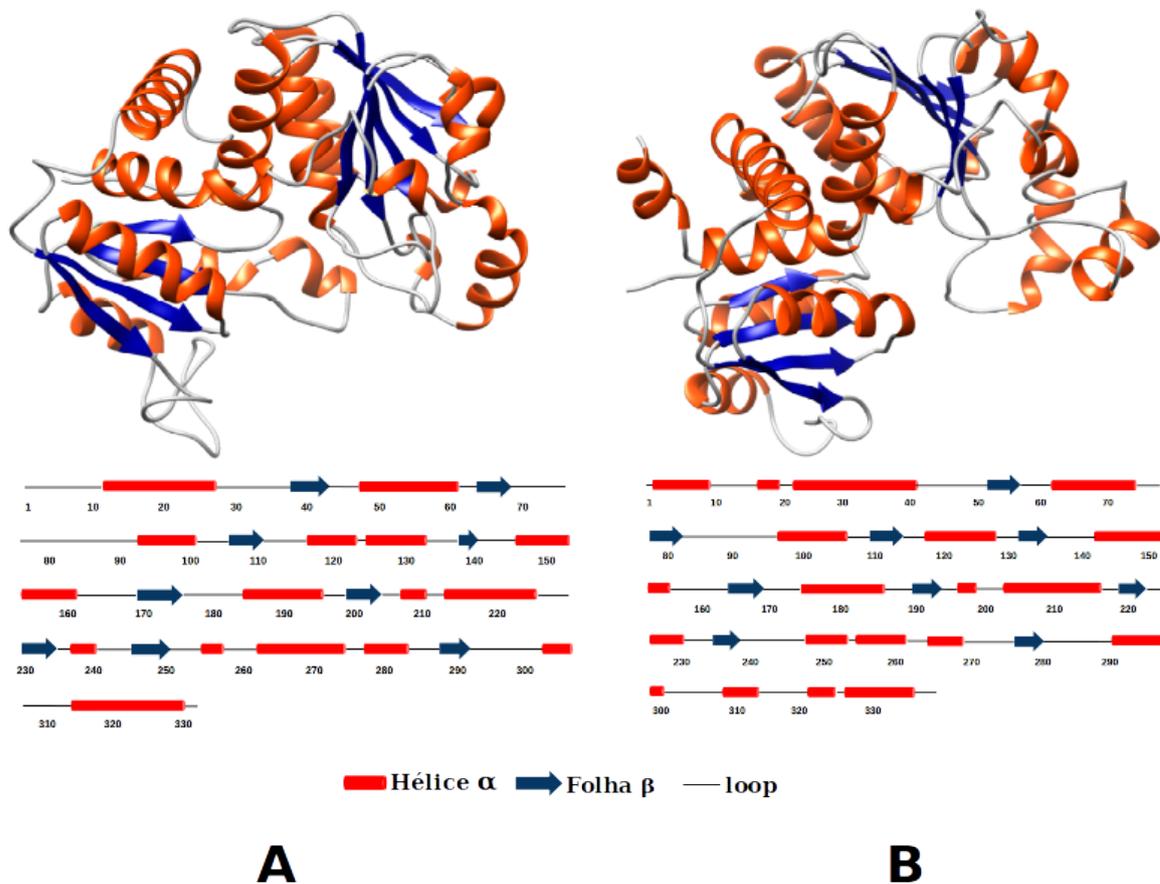


Figura 11 – Estrutura das proteínas: PDB 3KZN (PcN) em A e PDB 4JQO B.

Fonte: criado com UCSF Quimera^[32].

As proteínas AOTCase e OTCCase, que possuem função enzimática, têm como característica interessante sob o ponto de vista de estudo para PcN, como dito anteriormente, possuem o arranjo tridimensional muito similar e, além disso, catalizam reações semelhantes. No entanto, somente AOTCase possui em sua estrutura um nó^[33]. Estas enzimas possuem entre si 25% de identidade geral em sua estrutura (alinhamento de estruturas realizado com o VMD^[12]).

Abaixo, na figura 12 apresentamos a sobreposição das estruturas terciárias e a representação do percentual de identidade. Na parte A é possível verificar pela sobreposição como estas possuem arranjo tridimensional similar (azul para a PcN e laranja a proteína sem nó), e em B o destaque para a região do nó. Na parte C é apresentada o percentual de identidade em azul para as duas proteínas, isto é, os resíduos que são conservados em ambas.

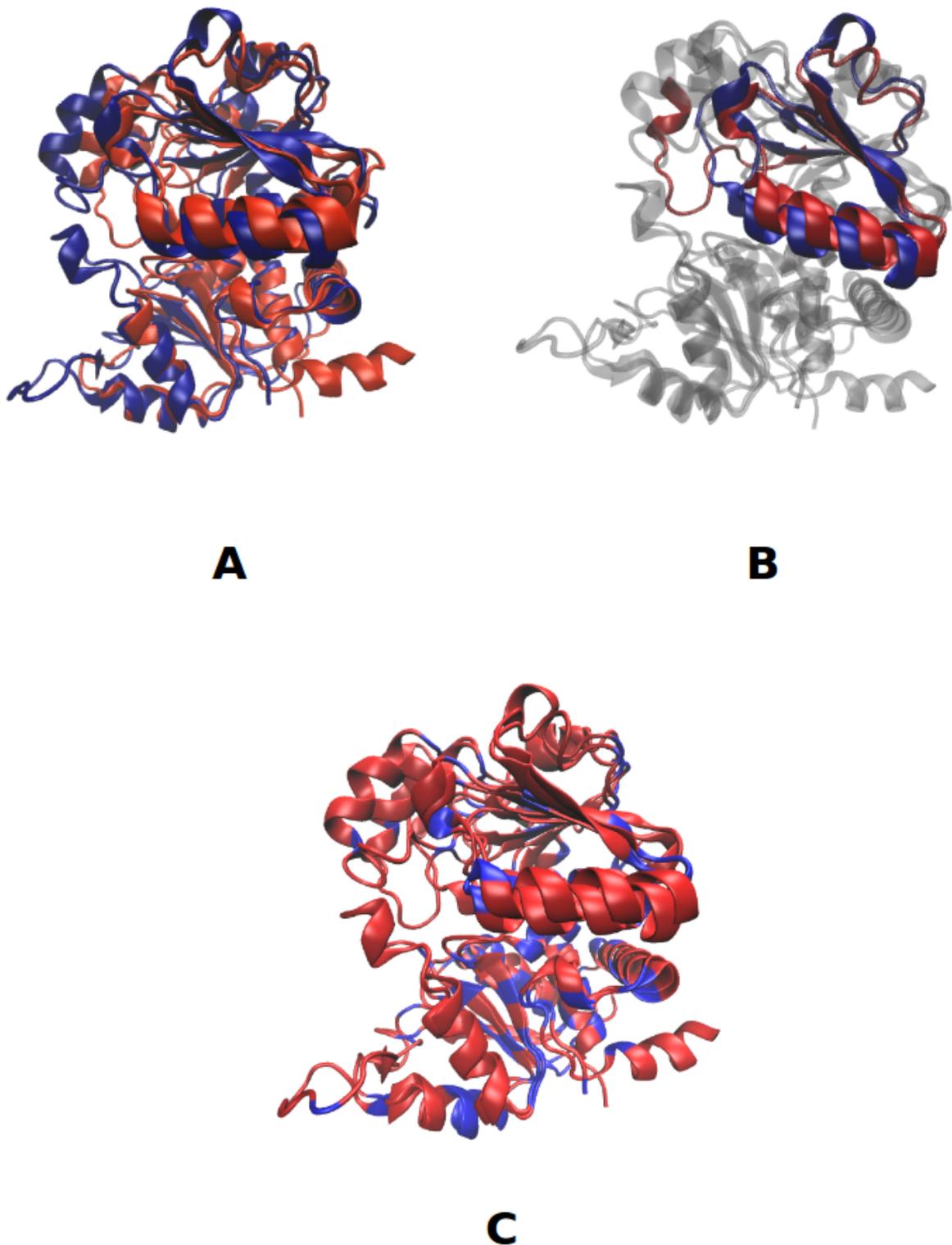


Figura 12 – Sobreposição das estruturas tridimensionais: **A** sobreposição de PDB 3KZN (PcN) em azul e PDB 4JQO em laranja; **B** destaque da região do nó e a região correspondente para a estrutura sem nó. Em **C** o alinhamento em relação a estrutura primária: em azul os resíduos conservados nas duas estruturas representam a baixa identidade (25%).

Fonte: criado com VMD.^[12]

As OTCs são enzimas chaves envolvidas no metabolismo da biossíntese de arginina em bactérias e plantas e no ciclo da ureia em animais^[14]. Dentro desta família, a ATO-Case é uma enzima essencial para a biossíntese de arginina em várias bactérias. Seu homólogo, OTCase está envolvida na degradação da arginina em algumas bactérias e participa do ciclo da ureia nos mamíferos^[60,61].

4.2 Identificação do Nó na PcN

Para visualização das características do nó presente na estrutura da proteína, de entrada PDB 3KZN foi utilizado o servidor para identificação de nós *KnotProt*^[16]. Com este é possível obter informação do tipo de nó, e seu detalhamento por meio de um diagrama de matriz denominado como uma "impressão digital" (*knotting fingerprint*) da PcN.

O *KnotProt* possui um banco de dados para PcN e também outros tipos de topologia. Desse modo é possível analisar uma estrutura já disponível no servidor ou submeter uma estrutura diretamente na sua página. Além disso, o servidor permite submeter trajetórias de modo a avaliar a presença de nó ao longo de uma simulação por dinâmica molecular^[16].

No diagrama de impressão digital obtido cada tipo de nó é representado por uma cor. Dessa forma, para uma proteína com nó do tipo 3_1 a cor apresentada é verde, 4_1 laranja, 5_2 violeta e 6_1 azul. A linha na diagonal do diagrama apresenta toda a extensão da proteína, sendo destacado a região correspondente ao nó como "*knot core*" e as regiões externas ao nó como "*knot tail*".

Na figura 13, é mostrado um exemplo de como o *KnotProt* representa as informações para uma PcN.

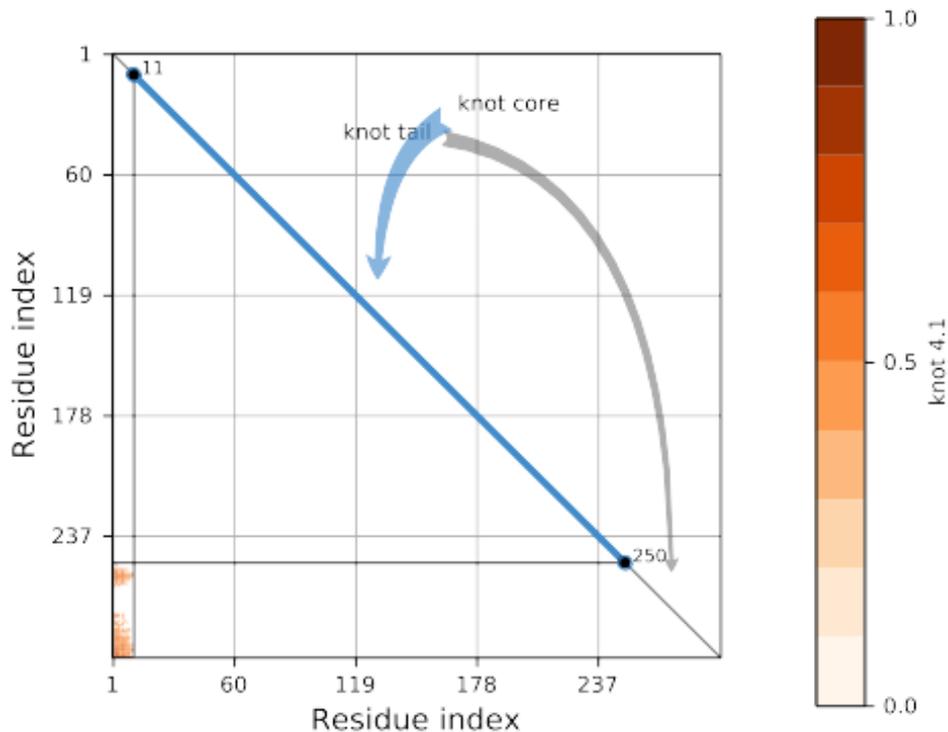


Figura 13 – Resultados obtido com o *KnotProt* para uma proteína com nó. O diagrama de refere a entrada PDB 5VIK.

Fonte: KnotProt^[16] .

Pelo diagrama de impressão digital apresentado na figura 13 para a proteína PDB 5VIK, o servidor fornece a informação de que a proteína tem nó do tipo 4_1 , cuja extensão vai do resíduo 11 até o 250 demarcado pela linha azul em destaque na diagonal (*knot core*). A região externa ao nó é representada na diagonal na cor cinza (*knot tail*). A proteína tem um total de 296 resíduos de aminoácidos.

4.3 Critérios de Simulação Molecular

As duas proteínas em estudo foram submetidas a simulações de Dinâmica Molecular com o software NAMD v2.12^[62] .

Inicialmente, utilizamos o Antechamber presente no pacote AmberTools18^[63] para parametrização do resíduo não padrão KCX, uma lisina acetilada presente na estrutura da PcN (PDB 3KZN). Em seguida, o servidor PropKa^{1[64]} foi utilizado para ajuste na protonação das cadeias laterais em pH 7.4.

¹<<http://server.poissonboltzmann.org/pdb2pqr>>

Os arquivos de topologia e coordenadas iniciais foram gerados com o software Leap do pacote AmberTools18^[63], utilizando o campo de força ff14SB^[65] para proteínas e modelo de solvente explícito TIP3P. Átomos de íons Na⁺ foram adicionados para neutralizar os sistemas (8 para PDB 4JQO e 7 para PDB 3KZN). Cada proteína foi confinada em uma caixa de solvatação retangular com 20Å de distância entre a proteína e as bordas.

As simulações foram realizadas em condições de contorno periódicas, com restrições de ligação H utilizando o algoritmo Shake^[66] e passo temporal de 2fs. O método Particle-Mesh Ewald (PME) foi usado para tratar as interações eletrostáticas de longo alcance com corte de 9Å.

As etapas da simulação foram:

Minimização: esta etapa foi realizada utilizando o algoritmo do gradiente conjugado inicialmente por 5000 passos mantendo fixa a proteína, em seguida 8000 passos de minimização foram realizados com todo o sistema sem restrição.

Aquecimento: após a minimização o sistema foi submetido a uma etapa de aquecimento gradual de modo a elevar a temperatura de 0 a 310K. Este procedimento foi realizado em três estágios:

Estágio 1: simulação realizada por 50ps com aquecimento gradual de 0 a 100K, impondo uma restrição harmônica com força contante de 30 kcal/mol/Å² para os átomos de C α e 20 kcal/mol/Å² para os átomos de C e N do *backbone* da proteína.

Estágio 2: simulação realizada por 100ps com aquecimento gradual de 100 a 200K, impondo restrição harmônica com constate de força de 20 kcal/mol/Å² para os átomos de C α e 5 kcal/mol/Å² para os átomos de C e N do *backbone* da proteína.

Estágio 3: simulação realizada por 200 ps com aquecimento gradual de 200 a 310K, impondo uma restrição harmônica com força contante de 10 kcal/mol/Å² para os átomos de C α e 2,5 kcal/mol/Å² para os átomos de C e N do backbone da proteína. Após este estágio final a temperatura do sistema foi mantida constante utilizando o termostato de Langevin^[67].

Pressurização: nesta etapa, onde a temperatura foi mantida constante em 310K, o sistema foi submetido a condição de pressão de 1atm, a qual foi mantida por meio do barostato de Langevin^[67]. Dois estágios foram utilizados nesta etapa:

Estágio 1: aplicando restrição harmônica com força constante de 5 kcal/mol/Å² para os átomos de C α e 1,25 kcal/mol/Å² para os átomos de C e N do *backbone* da proteína, o sistema foi simulado por 100ps.

Estágio 2: aplicando restrição harmônica com constante de força de 2,5 kcal/mol/Å² somente aos átomos de C α o sistema foi simulado por 200ps.

Equilibração: após a etapa de pressurização o sistema foi conduzido a uma etapa de equilibração por um total de 10ns, onde foram mantidas constantes a temperatura em 310 K e a pressão em 1 atm.

Produção: visando avaliar o comportamento dinâmico das estruturas utilizadas no trabalho, foram obtidas trajetórias de equilibração com total de 50ns nas temperaturas de 310K, 373K e 436K (cada temperatura foi realizada em triplicata).

4.4 Análises de Estabilidade

Para avaliar o comportamento das estruturas ao longo das simulações de dinâmica molecular realizadas foram utilizadas algumas métricas, descritas a seguir.

-Raíz do Desvio Quadrado Médio (RMSD, do inglês *Root Mean Square Deviation*)

A estabilidade das estruturas obtidas ao longo de uma simulação geralmente é avaliada pelo RMSD, onde se escolhe uma estrutura de referência e a partir desta, se calcula o desvio para cada uma das estruturas geradas em casa passo da trajetória obtida na simulação. Neste trabalho foi o utilizado o programa *cpptraj*, disponível no pacote do AmberTools18^[63]. Este programa oferece a possibilidade de análise de trajetórias obtidas em simulação empregando dinâmica molecular.

A equação para a determinação do RMSD é descrita como:

$$RMSD = \sqrt{\frac{\sum_{i=0}^N [m_i (X_i - Y_i)^2]}{M}} \quad (4.1)$$

Onde, N é o número de átomos, m_i é a massa do átomo i , X_i é o vetor das coordenadas obtidas para o átomo i , Y_i é o vetor de coordenadas da referência obtidas para o átomo i e M é a massa total.

Raíz da Flutuação Quadrática Média (RMSF, do inglês *Root Mean Square Fluctuation*)

O RMSF permite avaliar a flutuação das estruturas ao longo da trajetória em relação a posição da estrutura de referência. Os dados obtidos dessa métrica são utilizados para avaliar a flexibilidade de determinados grupo na molécula em simulações DM. Altos valores do RMSF indicam que houve grande oscilação para um determinado grupo de átomos, podendo ser utilizada para avaliar uma região específica da estrutura. Para a determinação do RMSF foi utilizado o *cpptraj* disponível no pacote do Ambertools18^[63].

Fração de Contatos Nativos (Q)

A coordenada Q , fração de contatos nativos, é dada pela razão entre o número dos contatos existentes em uma dada conformação e a estrutura nativa para uma determinada molécula^[68].

Os valores de Q variam entre 0 (quando a proteína perde sua estrutura nativa) e 1 (onde exibe a conformação nativa). Desse modo essa métrica pode ser utilizada para visualizar a diminuição de contatos nativos da estrutura durante a simulação, e com isso se pode visualizar contatos que sejam importantes para manter um determinado arranjo.

Neste trabalho a fração de contatos nativos (Q), foi calculada utilizando a biblioteca MDTraj^[69]. Esta biblioteca de acesso livre permite a análise de trajetórias de dinâmica molecular e tem como vantagem o fato de ser flexível e uso fácil.

4.5 Avaliação Conformacional por PCA e DCCM

Para avaliar o comportamento conformacional das proteínas utilizadas no estudo foram realizadas análises empregando Análise de Componentes Principais (PCA, sigla para o inglês *principal component analysis*) e Mapa dinâmico de Correlação Cruzada (DCCM, sigla para o inglês *dynamical cross-correlation map*), para o cálculo de ambas foi utilizada a biblioteca Bio3D^[70].

A PCA é empregada para avaliar mudanças ou tendências conformacionais da estrutura ao longo do tempo na trajetória de DM, de modo a reduzir o número de dimensões necessárias para a descrição da dinâmica proteica^[71]. Por meio das componentes principais (PCs) é possível descrever a variância máxima da distribuição das estruturas, e assim explorar o modo dominante do movimento molecular^[72].

O cálculo da PCA foi realizado com as coordenadas sobrepostas e alinhadas dos átomos $C\alpha$, empregadas para a diagonalização da matriz de covariância^[73]. Para o cálculo da matriz de covariância a fórmula utilizada foi:

$$C_{i,j} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle \quad (4.2)$$

Onde X_i e X_j representam respectivamente, as coordenadas atômicas de $C\alpha$ para o átomo i e j . O símbolo $\langle \rangle$ representa a média das coordenadas para a trajetória de simulação MD.

No cálculo DCCM são analisadas as correlações entre os resíduos da proteína utilizando as coordenadas do $C\alpha$ da trajetória DM. Dessa forma é possível avaliar as diferenças de correlação nas proteínas, por meio de análise visual de padrões nos gráficos^[72,74] e seu comportamento para diferentes variações de temperatura. A correlação entre os resíduos foi calculada pela equação:

$$C_{i,j} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{(\sqrt{\langle \Delta r_i^2 \rangle} \cdot \sqrt{\langle \Delta r_j^2 \rangle})} \quad (4.3)$$

Onde Δr_i e Δr_j correspondem aos deslocamentos da posição média para os átomos i e j no tempo, $\langle \rangle$ o tempo médio ao longo da trajetória. Valores de C_{ij} aparecem entre 1 e -1, de modo que valores positivos correspondem a movimentos de correlação e valores negativos movimentos de anticorrelação entre os resíduos^[72,75,76].

4.6 Cálculo de descritores de reatividade

Com o intuito de investigar características eletrônicas das estruturas foram realizados cálculos de descritores de reatividade utilizando o software PRIMoRDIA^[47]. Este software permite o cálculo de propriedades eletrônicas de macromoléculas graças a implementação de diversos descritores de reatividade globais e locais.

Ao todo 400 estruturas foram usadas nos cálculos dos descritores, sendo selecionadas com intervalos de 100 frames para as temperaturas 310K e 373K para as duas proteínas utilizadas no estudo.

Para obtenção das estruturas eletrônicas usadas no PRIMoRDIA, foram realizados cálculos *single-point* utilizando o pacote MOPAC^[77] com a opção do hamiltoniano semiempírico PM7^[78] e algoritmo de escalonamento linear MOZYME^[79] e modelo de solvente implícito COSMO.

Para investigar características da estrutura eletrônica, a avaliação de descritores de reatividade local foi direcionada para os descritores: nucleofilicidade, eletrofilicidade, netfilicidade, moleza e dureza locais, suscetibilidade radical, localização do orbital molecular, multifilicidade e densidade eletrônica. Todos esses avaliados em relação aos resíduos que são conservados para as duas proteínas na região do nó (Figura 14).

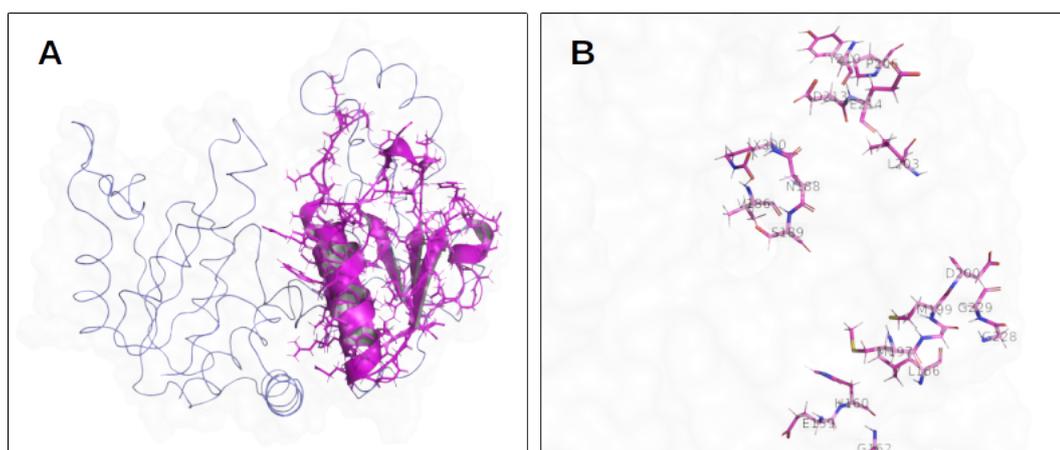


Figura 14 – PcN 3KZN: **A** Destaque para a região do nó em magenta. **B** Resíduos da região do nó usados para avaliação dos descritores.

Capítulo 5

Resultados e Discussão

5.1 Identificação do nó em 3KZN

As informações sobre a identificação e classificação do nó para a proteína de entrada PDB 3KZN, obtidas por meio do servidor KnotProt^[16] estão apresentadas na figura 15.

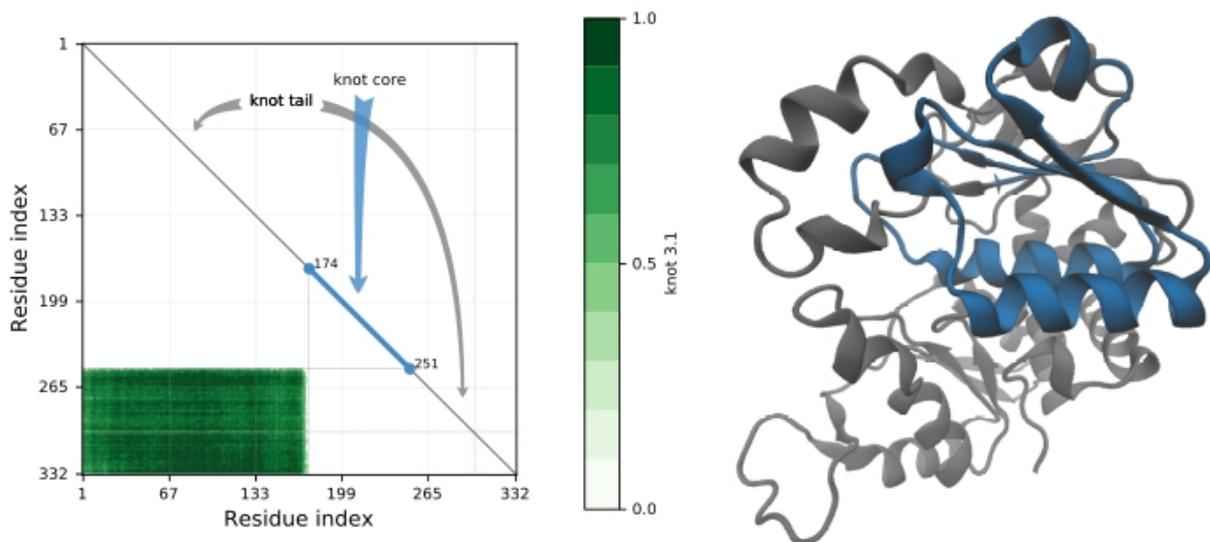


Figura 15 – Identificação do nó do tipo 3_1 na estrutura da proteína 3KZN. Na esquerda o diagrama representativo obtido com o *KnotProt*, a direita a representação tridimensional da estrutura com as cores correspondentes a informação do diagrama.

Pelas informações apresentadas no diagrama o nó identificado pelo servidor é do tipo 3_1 , que é acompanhado pela demarcação no diagrama na cor verde, cujo

tamanho total é de 78 resíduos, conforme destacado como *knot core* representado pelo segmento em azul na linha diagonal no diagrama da figura acima. O nó da estrutura está localizado entres os resíduos 174 e 251, estando situado a menor distância da extremidade C-terminal, 81 resíduos, e 173 resíduos para N-terminal, demarcado no diagrama pelos seguimentos de linha em cinza, para região denominada *knot tail*.

5.2 Análises de Estabilidade Estrutural

As trajetórias obtidas com 50ns de simulação foram analisadas por RMSD, RMSF e fração de contatos nativos (Q). Os dados apresentados nas figuras 16 e 17 são referentes a triplicatas em cada uma das temperaturas avaliadas.

Conforme se pode observar para as duas estruturas o tempo usado na simulação forneceu estruturas equilibradas tendo em vista que os valores de RMSD não apresentam picos ou grandes flutuações ao longo das trajetórias, variando em torno de 1Å ao longo do tempo simulado para as temperaturas de 310K e 373K.

Para a temperatura mais elevada, como se poderia esperar, os valores obtidos para 436K foram maiores uma vez que, elevando a temperatura, o sistema responde com possíveis variações. Embora para uma temperatura tão elevada os valores de RMSD tenham sido mais elevados, não houve um pico em destaque que demonstrasse desestabilização completa em algum momento da trajetória.

Em 310K é possível observar que para a região do nó, resíduos 174 ao 251, demarcado pelo retângulo em preto, os valores de flutuações não são muito elevados para a PcN (PDB 3KZN em azul). Já para a proteína sem nó (PDB 4JQO em verde), considerando a mesma região se observa um valor elevado de flutuação de 2,5Å para o resíduo 243 (GLY243), o qual está situado numa região de *loop*. Apesar da ausência de flutuações significativas na região do nó, ainda é possível observar grandes valores de RMSF na PcN fora dessa região. Por exemplo, uma região de grandes flutuações entre os resíduos 86 e 88 (ASP86, GLY87, ASP88), todos eles numa região de *loop*, o que justifica uma maior flexibilidade em comparação com os valores destacados na figura 16.

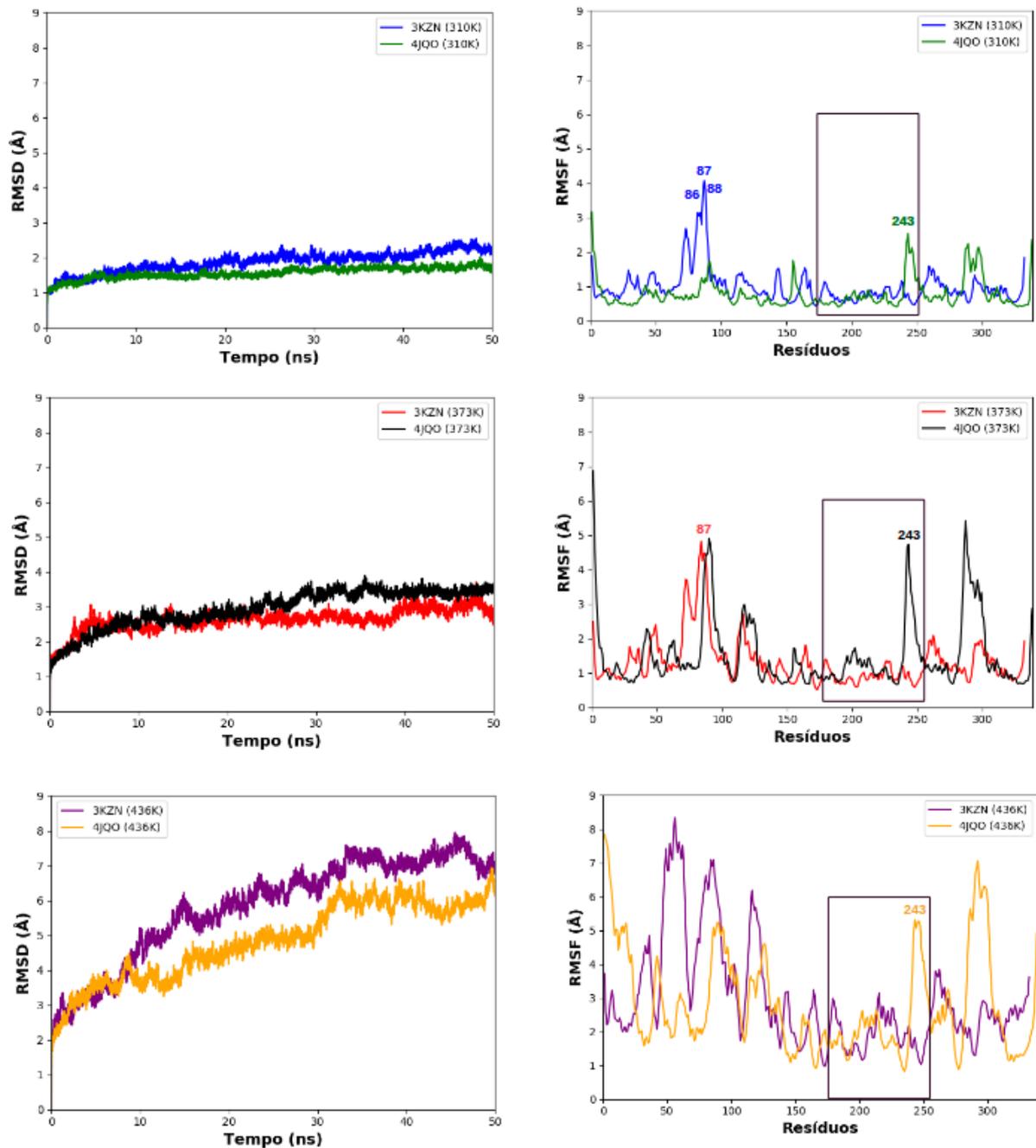


Figura 16 – RMSD e RMSF (50ns) para as proteínas PDB 3KZN (PcN) e 4JQO. Valores em azul e verde para 310K, vermelho e preto para 373K e roxo e amarelo para 436K. O retângulo demarca a região do nó na PcN. Os valores são referentes a triplicata.

Na temperatura de 373K, podemos perceber que, mesmo com a elevação da temperatura, a região do nó na PcN se manteve sem grandes flutuações, enquanto o valor de RMSF para o resíduo 243 (GLY243), localizado na mesma região, sofreu aumento (4,7Å) na proteína sem nó.

Além disso, fora da região do nó observamos valores de flutuação mais elevados para a proteína sem nó (PDB 4JQO) e uma flutuação acentuada para a PcN (PDB 3KZN) para o resíduo 87 que está fora do nó e compõe um *loop*. Assim, observando o comportamento geral, a estrutura que possui nó apresentou menores flutuações ao longo da simulação quando submetida a temperatura de 373K.

Como era esperado ao impor um aumento significativo de temperatura ao sistema, 436K, houve maior flutuação de modo geral em ambas as estruturas. O comportamento de RMSD foi bem similar para as duas proteínas.

Em relação aos valores de RMSF, As maiores flutuações permaneceram fora da região do nó. O resíduo 243 (GLY243 destacado) apresentou uma flutuação de 5,3Å, para a proteína sem nó (linha laranja). As regiões fora do nó apresentaram maior flutuação para a PcN abaixo da região que contém o nó (entre os resíduos 1 e 173) e para a proteína sem nó, acima do região correspondente ao nó (resíduos acima de 251).

A fração de contatos nativos (Q) para as duas estruturas em estudo são apresentadas na figura 17. Comparando os perfis para as proteínas se verifica um comportamento semelhante em 310K. As trajetórias azul e verde não apresentaram diminuição significativa dos contatos nativos, tendo valores acima de 0,95 ao longo de toda a trajetória.

Em 373K, ambas as proteínas mostram uma pequena diminuição da fração Q. Embora a curva em vermelho correspondente a PcN mostre uma diminuição maior, são valores ainda muito próximos da curva em preto para a proteína sem nó.

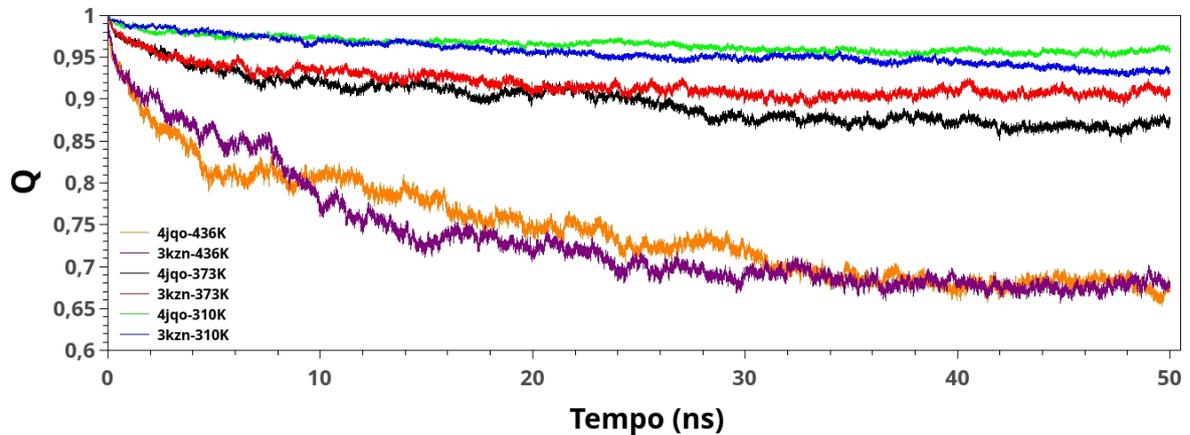


Figura 17 – Fração de contatos nativos (Q) para 50ns de simulações de DM. A PcN é apresentada nas cores azul, vermelha e roxa para 310K, 373K, 436K, respectivamente. A proteína sem nó está representada em verde, preto e laranja para essas temperaturas. Valores referentes a triplicata

Desse modo ainda não foi possível afirmar que há uma tendência da PcN para maior termoestabilidade, como o observado por Xu e colaboradores^[3], ao avaliarem uma proteína consideravelmente menor que as utilizadas nesse estudo. Vale ressaltar, que no referido trabalho a proteína foi comparada com ela mesma após a modificação espacial de um dos cruzamentos da cadeia que o caracterizava. Neste estudo foram avaliadas duas estruturas sem tal modificação.

Com o intuito de avaliar a resistência à diminuição, a temperatura foi elevada para 436K. As curvas em roxo e amarelo, representam os valores obtidos para a PcN e a proteína sem nó, respectivamente. É possível observar na figura 17 que houve uma diminuição significativa de contatos nativos para ambas as estruturas. No entanto, para o tempo simulado os perfis não mostraram comportamento distinto entre as duas estruturas estudadas.

5.3 Análises do Comportamento Conformacional por PCA e DCCM

Além das análises convencionais realizadas (RMSD, RMSF, fração de contatos nativos), apresentamos também dados da avaliação dinâmica das estruturas por meio da Análise de Componentes Principais (PCA) e Mapa Dinâmico de Correlação Cruzada

(DCCM).

Por meio da PCA é possível analisar o comportamento conformacional das estruturas ao longo das trajetórias DM, permitindo conhecer importantes movimentos biomoleculares^[80]. Normalmente são apresentadas as duas primeiras componentes principais (PCs), que juntas são capazes de explicar a maior variância das conformações visitadas durante a simulação^[72].

Os dados referentes a cada temperatura de estudo, para cada proteína são apresentados mostrando as contribuições dos resíduos para a variância dos dados. Nos gráficos da PCA, a variância dos dados, ou seja, a variação das conformações capturadas durante o tempo de simulação, é representada como pontos indo do azul ao vermelho em ordem do tempo das trajetórias obtidas.

Dessa forma, o aglomerado dos pontos de cada cor representa mudanças no estados conformacionais. Em azul e vermelho estados conformacionais definidos, e em branco estados intermediários^[81,82].

A Figura 18 apresenta os dados para a proteína com nó em 310K. Nessa temperatura a variância total dos dados apresentados pelas duas primeiras PCs (PC1xPC2) foi de 45%.

A distribuição dos dados indo do vermelho ao azul aparecem muito agregadas para a PcN na menor temperatura de estudo, o que indica alta estabilidade da estrutura ao longo dos 50 ns obtidos na trajetória (Figura18: A).

Pelo gráfico dos resíduos versus PCs (Figura18: B), foi possível observar que não há grande destaque de nenhum resíduo em relação ao que aparece em PC1 (linha em preto) e PC2 (linha em azul). Dessa forma, o comportamento observado para as conformações é resultado de todos os resíduos, não sofrendo impacto de uma região específica da estrutura.

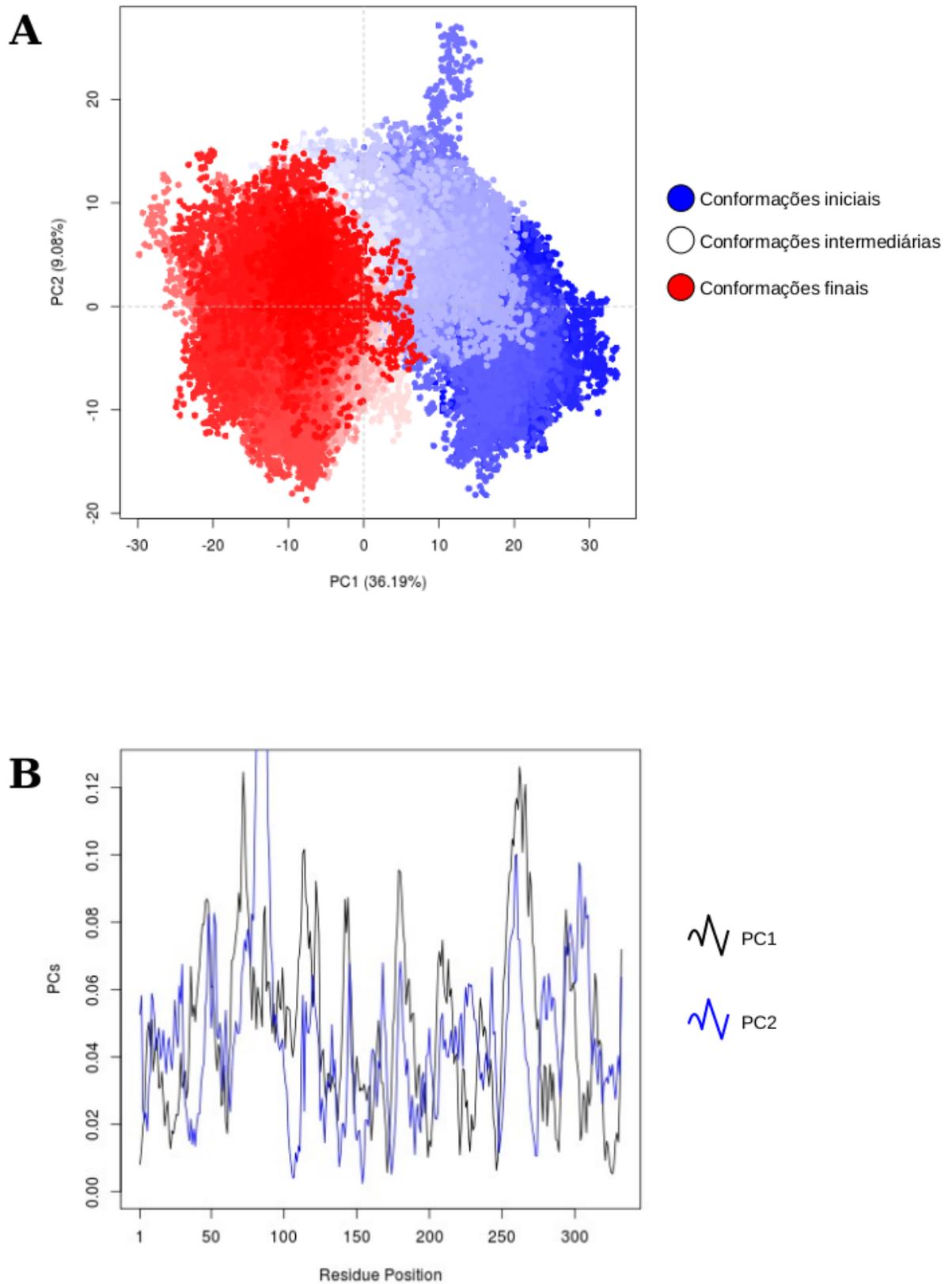


Figura 18 – PCA para a PcN (PDB 3KZN) em 310K. **A** variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

Para a segunda temperatura de estudo, 373K, o total de variância explicada dos dados foi de 39% considerando as duas primeiras PCs. Os dados são apresentados na Figura 19.

Com a elevação da temperatura os dados obtidos mostram que as conformações aparecem agregadas, reforçando que a PcN apresenta estabilidade. Na parte **B** da Figura 19, o gráfico resíduos versus PCs aponta destaque para os resíduos do 70 ao 90 para PC1 e PC2, o que aponta essa região com a maior contribuição para a variância observada na PCA.

Embora seja observado maior contribuição para as duas primeiras PCs apresentadas na PCA dos resíduos da região entre 70 e 90, estes não geraram impacto na distribuição das conformações, ou seja, ainda há grande densidade como é exposto pela alta concentração de conformações vermelho e azul, o que aponta para alta estabilidade mesmo para uma temperatura de 373K.

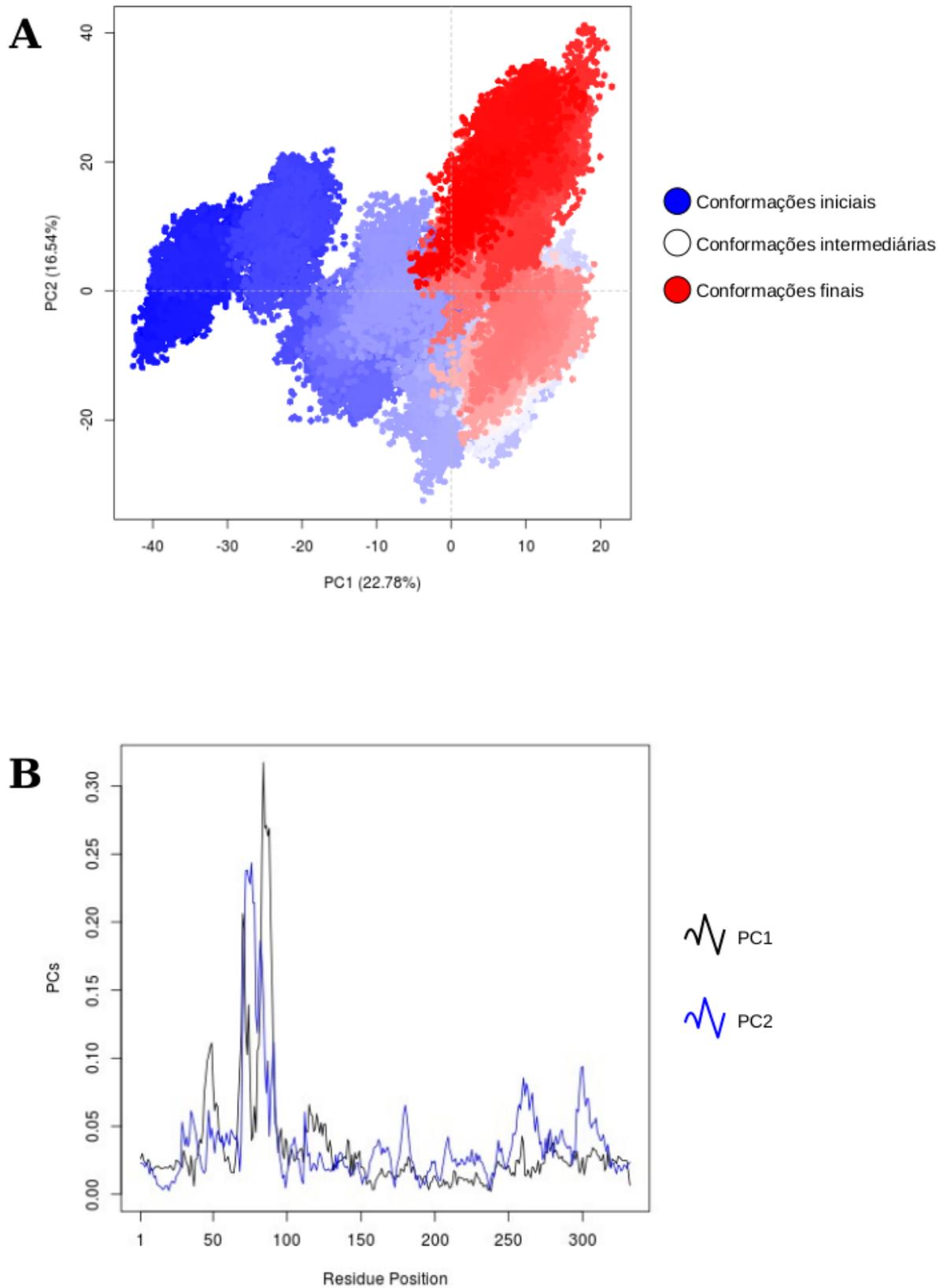


Figura 19 – PCA para a PcN (PDB 3KZN) em 373K. **A** variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

Buscando avaliar a estabilidade conformacional da proteína com nó, a temperatura de estudo foi elevada para 436K. Nessa temperatura, o total de variância explicada pelas duas primeiras PCs foi de 54%. Os dados são apresentados na Figura 20.

Para a temperatura de estudo mais elevada houve uma mudança em relação as demais como é possível observar na parte **A** da Figura 20. Ao longo da distribuição dos dados na PCA nota-se maior dispersão do vermelho ao azul, o que aponta que para essa temperatura a proteína perdeu parte da sua estabilidade. O que é condizente ao observado em outros estudos, onde a estabilidade de proteína com nó foi avaliada com temperatura muito elevada^[3].

Pelo gráfico de PCs versus resíduos mostrado em **B**, há contribuição de outros resíduos em geral para PC1 e PC2 quando em comparação as outras duas temperaturas do estudo. No entanto, a região de maior influência para a variância observada na PCA está entre os resíduos 50 e 70 para PC1 e 60 e 80 para PC2, o que fica aproximadamente em torno das maiores contribuições observadas nas outras temperaturas.

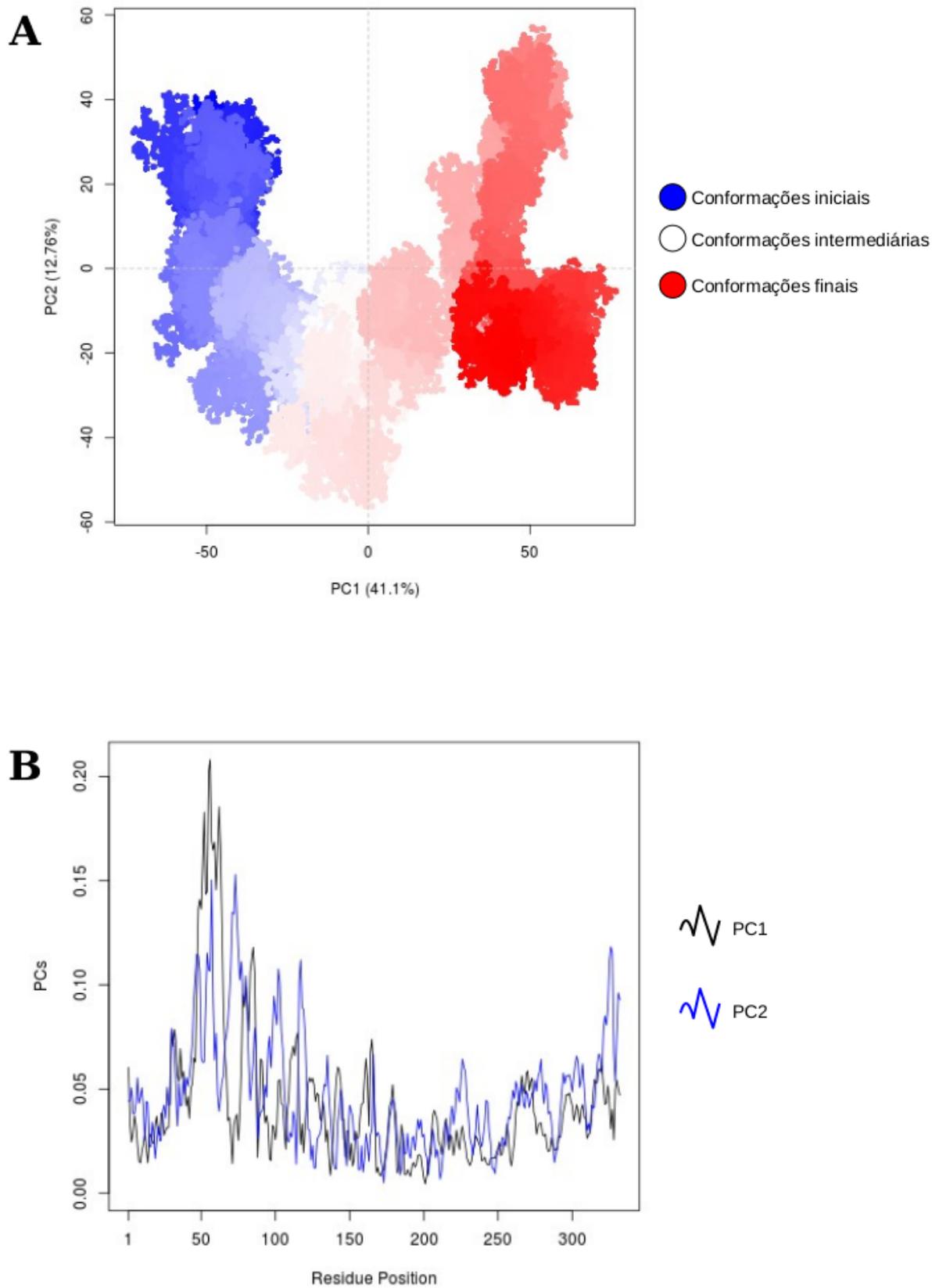


Figura 20 – PCA para a PcN (PDB 3KZN) em 436K. **A** variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

A PCA para a proteína sem nó (PDB 4JQO) apresenta comportamento bastante distinto em relação a PcN para a dispersão das conformações.

Em 310K, dados apresentados na Figura 21 o total de variância explicada dos dados foi de 37%.

Diferentemente do que se observa para a PcN na Figura 18, há uma concentração de conformações intermediárias (pontos brancos), o que aponta para menor estabilidade da proteína sem nós. Essas conformações em branco indicam que a proteína percorre uma maior mudança conformacional durante a trajetória nessa temperatura, distinto do comportamento para a proteína com nó, que tem maioria conformações em vermelho e azul bastante agregadas indicando menores mudanças conformacionais ao longo da trajetória simulada.

Nesta temperatura a amplitude de contribuição dos resíduos para as PCs é consideravelmente maior do que para a PcN. Além disso, vários resíduos se destacam com maior contribuição para PC1 (linha em preto) e PC2 (linha em azul) como se pode observar no gráfico PCs versus resíduos (Figura 21: **B**).

Um ponto interessante é o destaque do resíduo próximo a 250 que está na região correspondente ao nó na PcN, o que não é observado na parte **B** da figura 18 para a temperatura de 310K. O que corrobora com o observado para o RMSF, onde o resíduo 243 mostrou ter a maior flexibilidade.

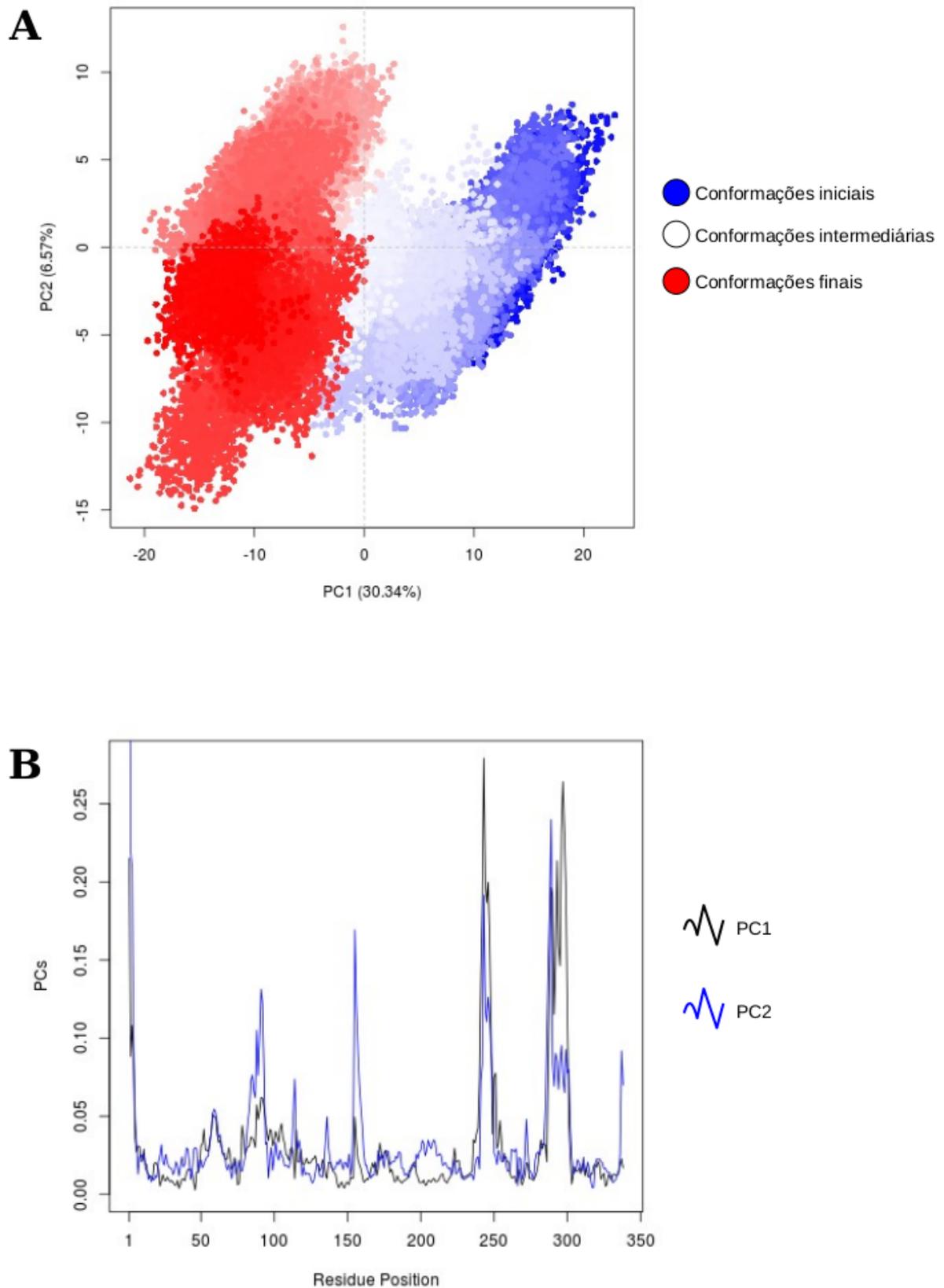


Figura 21 – PCA para a proteína sem nó (PDB 4JQO) em 310K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

A PCA com os dados referentes a proteína sem nó na temperatura de simulação de 373K são apresentados na Figura 22. Nesta temperatura as duas primeiras PCs foram capazes de explicar também 37% da variância dos dados.

Em 373K a PCA, embora tenha apresentado elevada densidade de conformações em vermelho e azul, similar ao que foi observado nessa temperatura para a estrutura com nó, apresenta também grande quantidade de conformações representadas na cor branca, o que indica que para essa temperatura a proteína sem nó apresenta maiores variações na conformação.

Em relação as contribuições dos resíduos para as PCs, apresentado na parte **B** da Figura 22, surgem diferenças de contribuição. Para PC1 (linha em preto) a maior contribuição aparece com o resíduo em torno de 300, o que não foi observado para a PcN. PC2 (em azul) apresentou dois picos significativos, o maior em torno do resíduo 90 e o menor próximo ao resíduo 250.

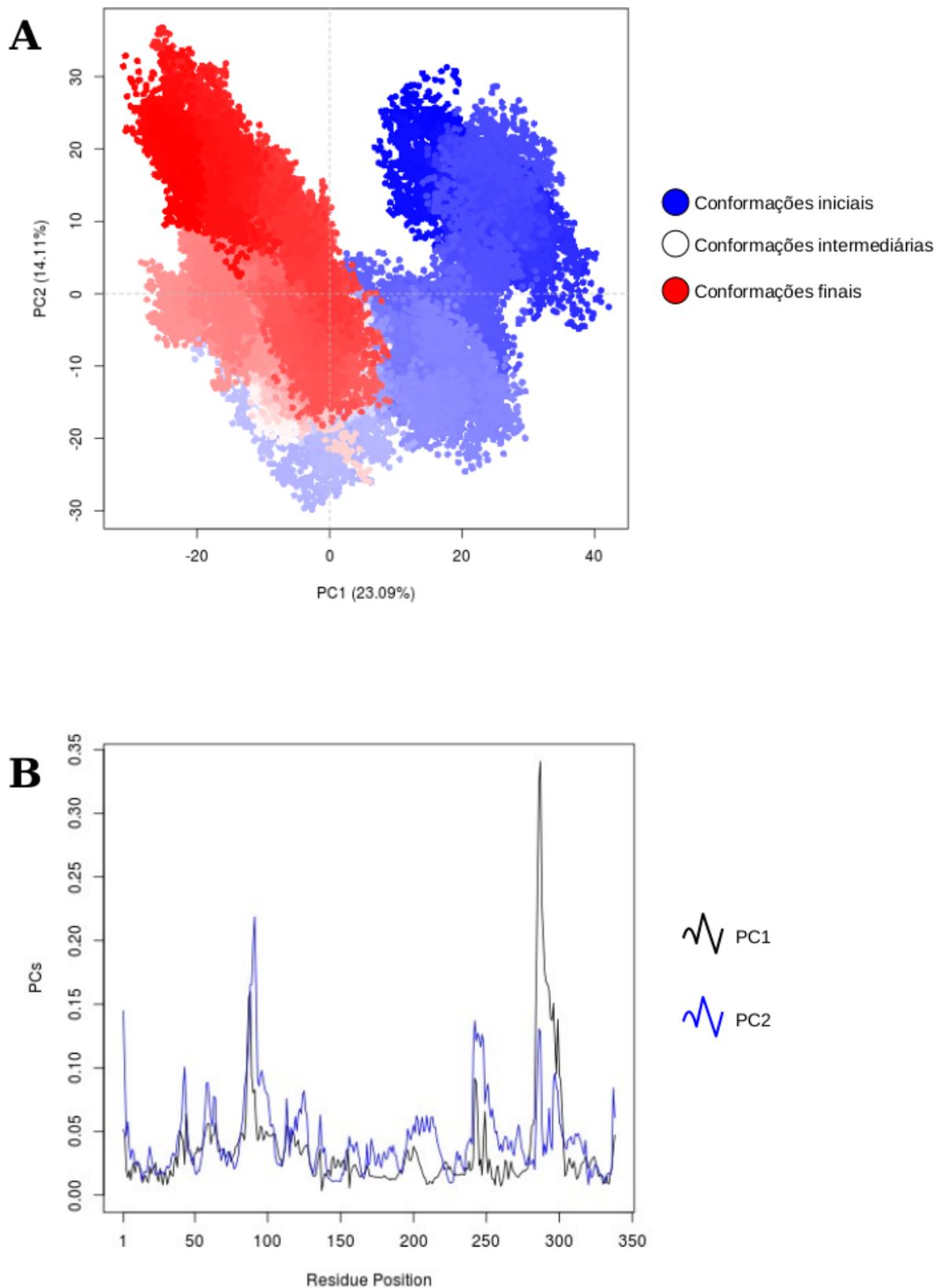


Figura 22 – PCA para a proteína sem nó (PDB 4JQO) em 373K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

Os dados para a análise de componentes principais para a proteína sem nós em 436K, são apresentados na Figura 23. O total de variância explicada dos dados com as duas primeiras PCs foi de 43%.

Para a temperatura de 436K a distribuição das conformações ao longo de PC1 é majoritariamente em vermelho (Figura 23: A). Nesta temperatura, essas conformações aparecem consideravelmente em maior quantidade que em comparação a PcN (Figura 20: A). Há grupos de conformações em vermelho entre 40 e 60 na PCA, um menor entre 20 e 40, e ainda outro aproximadamente entre -10 e 10. O que indica grupos de conformações variáveis para a estrutura, indicando que nesta temperatura houve maior variação das conformações.

Os resíduos de maior destaque para as PCs são os mesmos observados em 373K, no entanto, PC2 em azul traz os maiores picos em torno de 300 e logo abaixo de 250. Sendo estas as regiões de resíduos que mais tiveram contribuição para a variância observada na PCA.

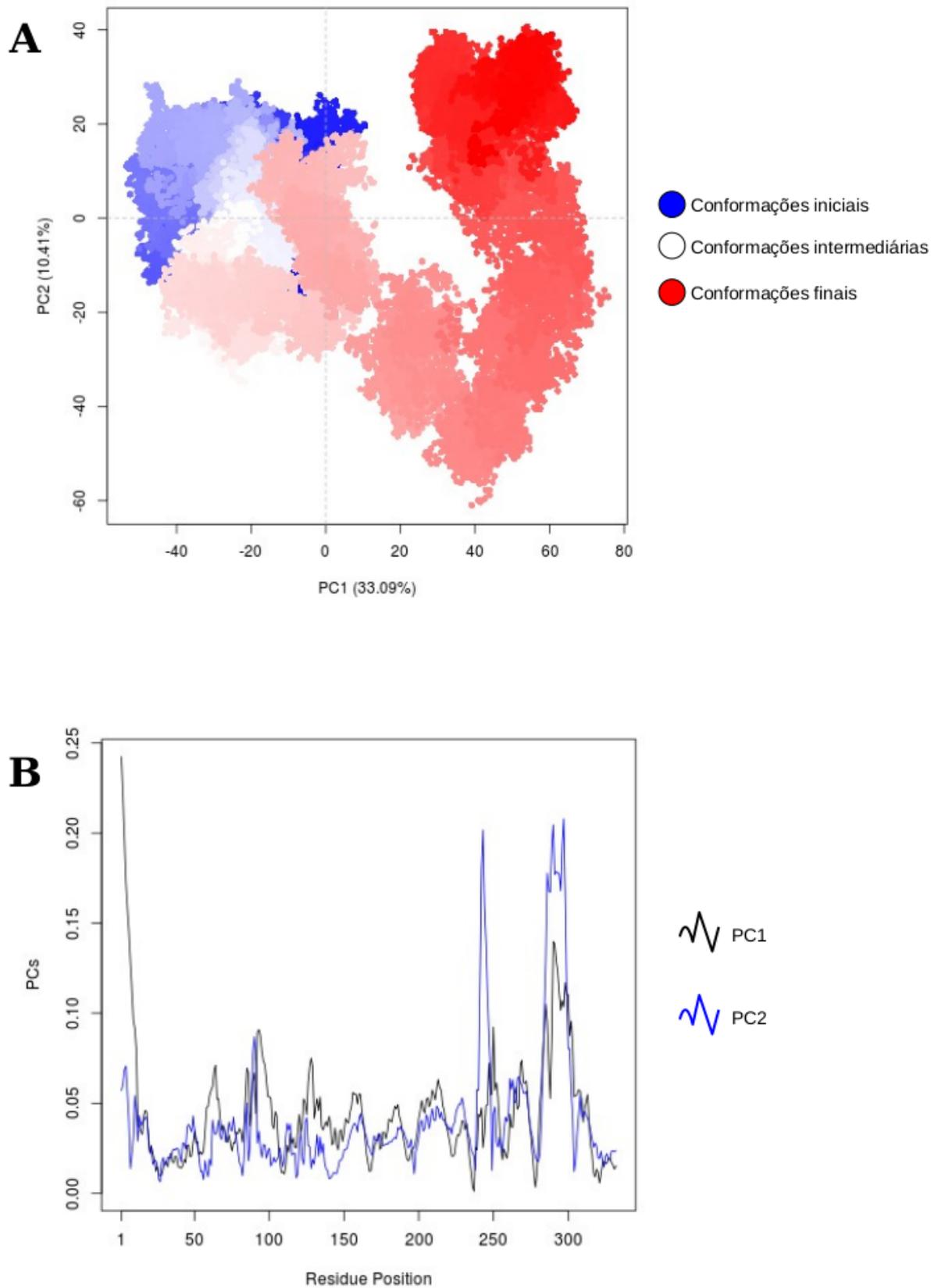


Figura 23 – PCA para a proteína sem nó (PDB 4JQO) em 436K. A variância dos dados (PC1xPC2), pontos em azul, branco e vermelho seguem a ordem do tempo na simulação. Em **B** as contribuições dos resíduos para as duas primeiras PCs.

O comportamento de correlação de movimento dos resíduos para as duas proteínas são apresentados nas figuras 24 e 25 para a proteína sem nó e a PcN, respectivamente. Os comportamentos de ambas as proteínas em toda a faixa de temperatura são tão distintos que mesmo uma inspeção visual superficial no gráfico é suficiente para percebê-lo.

Quanto maior a cor ciano (valores positivos), mais forte é a correlação no movimento dos resíduos expressa no gráfico DCCM, mostrando que os resíduos se movem na mesma direção. Caso contrário, quando prevalece a cor rosa (valores negativos) os movimentos ocorrem no sentido oposto, apresentando anticorrelação. Os retângulos na figura delimitam as regiões de mudança significativa observada com o aumento da temperatura.

Na Figura 24, referente a proteína sem nó, são apresentados os gráficos por temperatura: A em 310K, B em 373K e C para 436K.

Em 310K, observamos que há pouca correlação para a proteína sem nó. Existe apenas pequenas áreas em rosa claro no gráfico apresentado na parte **A**, mostrando movimentos de anticorrelação (regiões entre 240 e 250 e entre 280 e 300 com a região de 50 a 120).

Com o aumento da temperatura para 373K, a proteína sem nó apresentou um aumento nos movimentos como se pode observar na parte **B** da figura 24. Há um aumento considerável nos movimentos de anticorrelação no gráfico mostrado em **B**, como se pode notar pelo aumento da área rosa analisando a região dos resíduos 275 a 315 com a região entre 50 e 120. A correlação (em ciano) aparece em algumas regiões como o aumento observado na região acima de 315 com a região entre 145 a 155.

O comportamento em 436K para a proteína sem nó, como pode ser visto em **C**, destaca principalmente pelo surgimento de movimentos de correlação na região composta por resíduos entre 153 e 225 com a região entre 60 e 80, o que em temperaturas mais baixas não foi significativo. E, de modo geral, também houve aumento na anticorrelação, com destaque para a região entre 135 e 225 com a região entre 90 e 130.

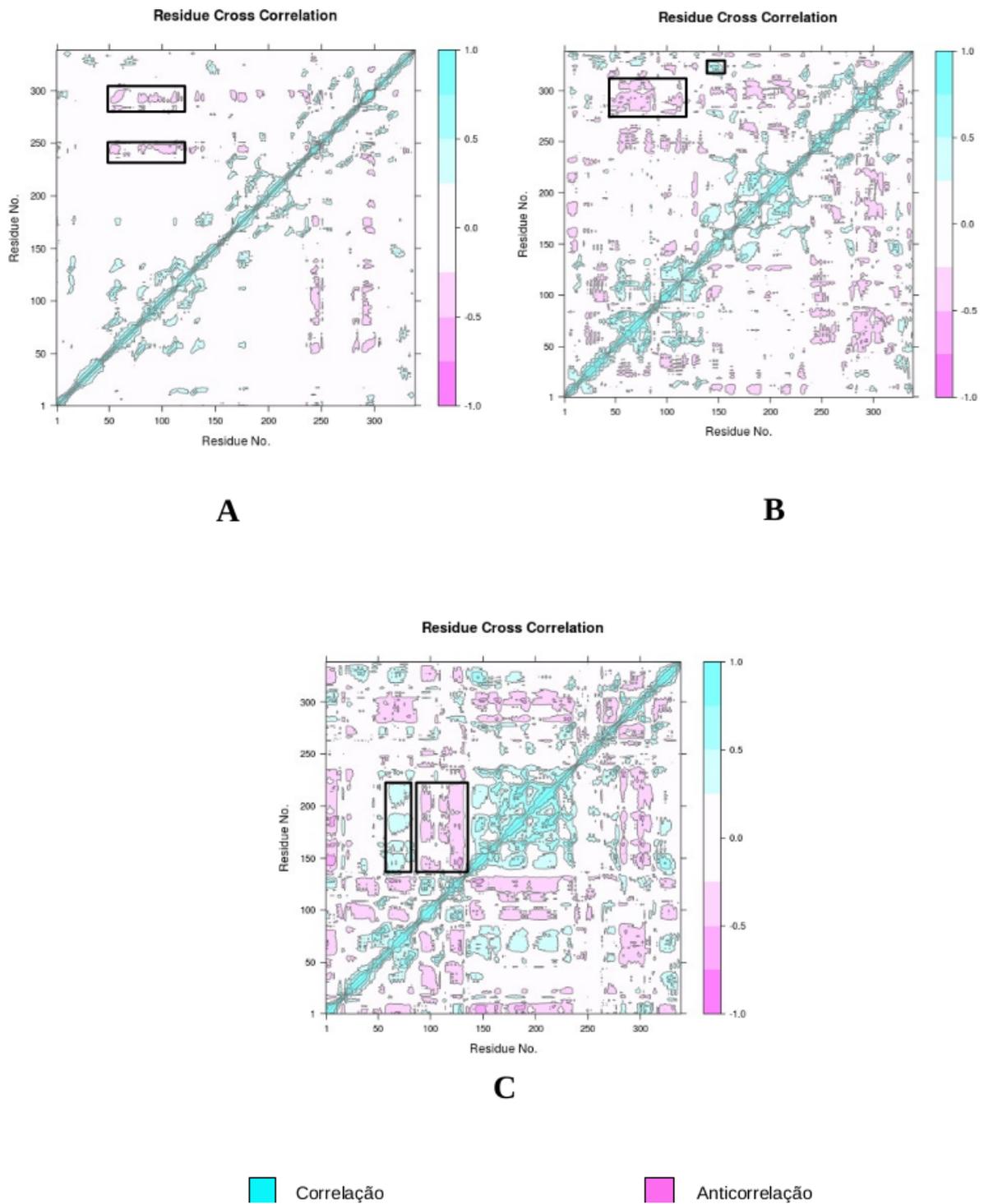


Figura 24 – DCCM para a proteína sem nó (PDB 4JQO). Os dados são apresentados por temperatura de estudo: **A** 310K, **B** 373K e **C** 436K. Os retângulos destacam as áreas de mudança significativa.

Considerando o comportamento total da proteína sem nó referente a elevação de temperatura realizada no estudo, se pode verificar que essa foi suscetível a mudança significativa nos movimentos de correlação e anticorrelação. Este comportamento

demonstra que a proteína tem comportamento dinâmico bastante variável a medida que é submetida a variações de temperatura.

Na Figura 25 são apresentados os gráficos DCCM para a PcN por temperatura: A em 310K, B em 373K e C para 436K.

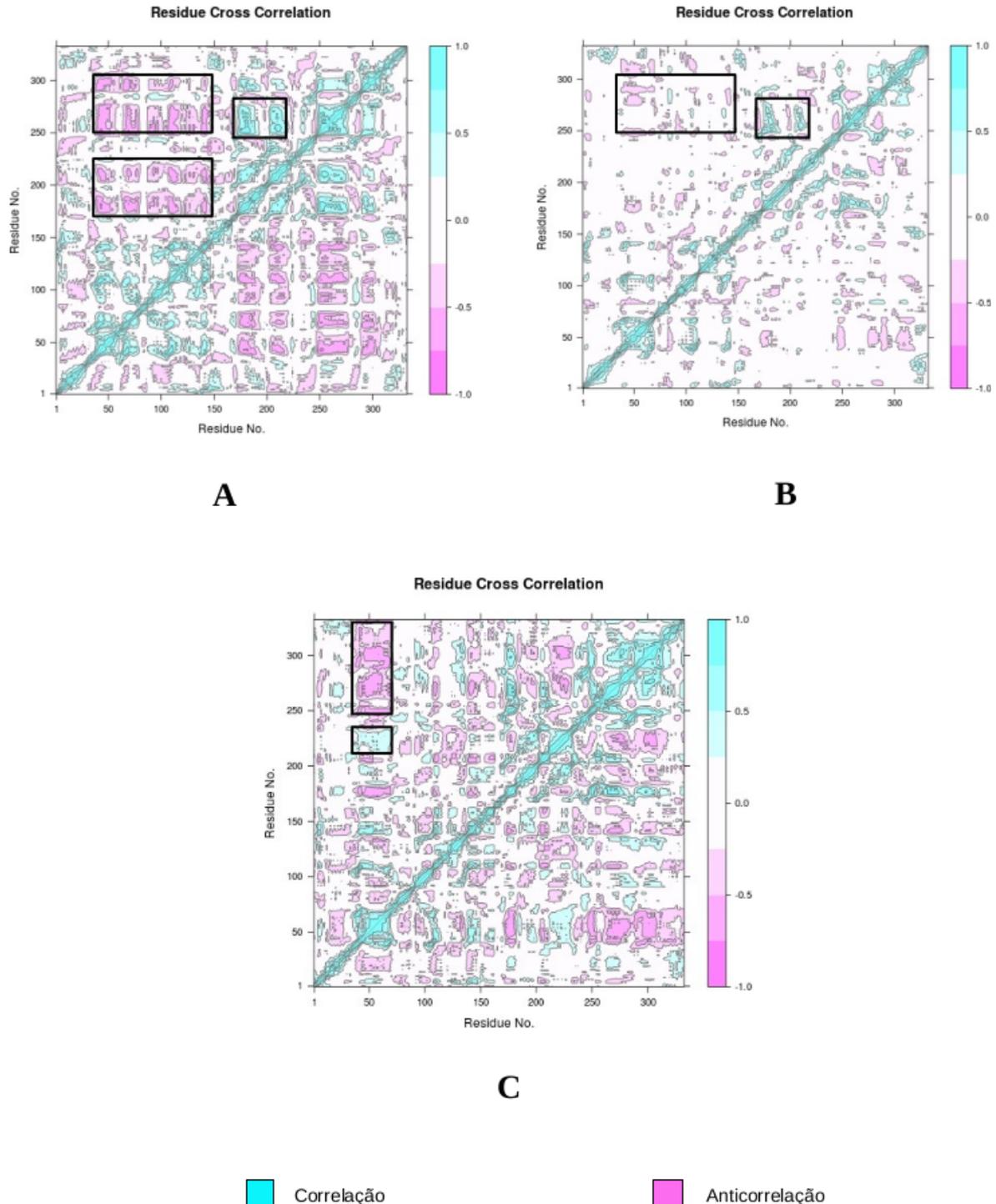


Figura 25 – DCCM para a proteína com nó (PcN) (PDB 3KZN). Os dados são apresentados por temperatura de estudo: **A** 310K, **B** 373K e **C** 436K. Os retângulos destacam as áreas de mudança significativa.

Para a temperatura mais baixa utilizada no estudo, 310K o gráfico DCCM na parte **A** da Figura 25 mostra que a PcN teve comportamento bastante distinto da proteína sem nó. Há grandes áreas referentes aos movimentos de correlação e anticorrelação dos resíduos, áreas destacadas nos retângulos na figura.

É possível notar regiões de coloração rosa em tom forte (valores próximo a -1), apontando para elevada anticorrelação (isto é, regiões de movimentos opostos) para as regiões de 175 a 220 e 250 a 300 com a região entre os resíduos 45 e 150. Há também para a PcN alta correlação (em ciano) entre a região de resíduos 250 e 275 com a região entre 175 e 225 que está inserida no nó.

Ao elevar a temperatura para 373K, Figura 25: **B**, notamos comportamento distinto: há diminuição dos movimentos. Esse comportamento é observado como resposta ao aquecimento da estrutura.

Em 373K houve uma diminuição acentuada dos movimentos de correlação e anticorrelação. Por exemplo, a anticorrelação da região 250 a 300 com a região de 45 a 150 é notavelmente reduzida. A região do nó tem o movimento de correlação observado entre as regiões 250 e 275 com a região 175 a 225, marcado em ciano, diminuído com a temperatura mais alta, mas visualmente ocorre menor redução quando em comparação a redução ocorrida para as regiões de anticorrelação citadas.

Na parte **C** da Figura 25, referente a temperatura de 436K é possível notar um aumento significativo dos movimentos de anticorrelação. Em destaque entre o segmento de resíduos 250-332 com a região entre 40 e 70. O movimento de correlação também aumentou em geral, por exemplo, a correlação entre os segmentos 215-230 e 40-70 apresentou um aumento considerável quando comparado a temperaturas mais baixas.

Ao avaliar as três temperaturas, é observado que a PcN, considerando todas as temperaturas teve a menor variação nos movimentos de correlação e anticorrelação. O que aponta para maior estabilidade em relação a proteína sem nó que teve maior modificação ao longo das temperaturas avaliadas.

Os dados de PCA e DCCM mostraram maior resistência a mudanças conformacionais para a PcN (PDB 3KZN) em comparação com a proteína sem nó (PDB 4JQO). O presente estudo mostrou que a PcN é mais estável nas condições de temperatura consideradas. Este resultado concorda com o obtido por Xu e colaboradores^[15], quando analisaram a importância do nó para uma proteína e observaram estabilidade redu-

zida para a estrutura sem nó. No estudo experimental realizado por Sirarimoju e colaboradores^[14], as proteínas nodosas também foram consideradas mais resistentes às mudanças impostas aos sistemas.

5.4 Descritores de Reatividade

Para os resíduos idênticos nas duas proteínas na região do nó, foram calculados os descritores de reatividade (DR) e analisadas sua distribuição de valores para encontrar padrões nas duas estruturas. A figura 26 apresenta o resumo da análise dos descritores calculados.

A densidade eletrônica, nucleofilicidade e a localização da banda orbital molecular são os descritores que apresentaram a mesma distribuição de valores para as duas proteínas.

Uma mudança sensível na densidade eletrônica dos resíduos é esperada quando eles participam de uma reação química, com a ocorrência de quebra e formação de ligações ou transferência de carga. Essas flutuações estruturais observadas na amostragem de dinâmica molecular com apenas campos de força clássicos não devem produzir mudanças tão sensíveis. Mesmo assim, o perfil de distribuição figura 26: J apresentado para esta propriedade mostra a similaridade química dos resíduos utilizados na análise dessas duas proteínas na região do nó.

Por outro lado, o descritor de eletrofilicidade mostra uma clara diferença no padrão de distribuição de seus valores entre os dois tipos de proteínas (Figura 26: B), mostrando que a proteína sem nó apresenta resíduos na região equivalente ao nó com maior reatividade, em particular acomodando a densidade eletrônica e agindo como um eletrófilo.

Essa diferença também foi encontrada em outros dois descritores de interações do tipo mole-mole, a moleza local (Figura 26: D) e a multifilicidade (Figura 26: H), mostrando como a proteína sem nó é mais eletrofílica e os valores do descritor apresentam um pico enquanto que para a PcN estão em torno do valor zero, mostrando sua menor reatividade.

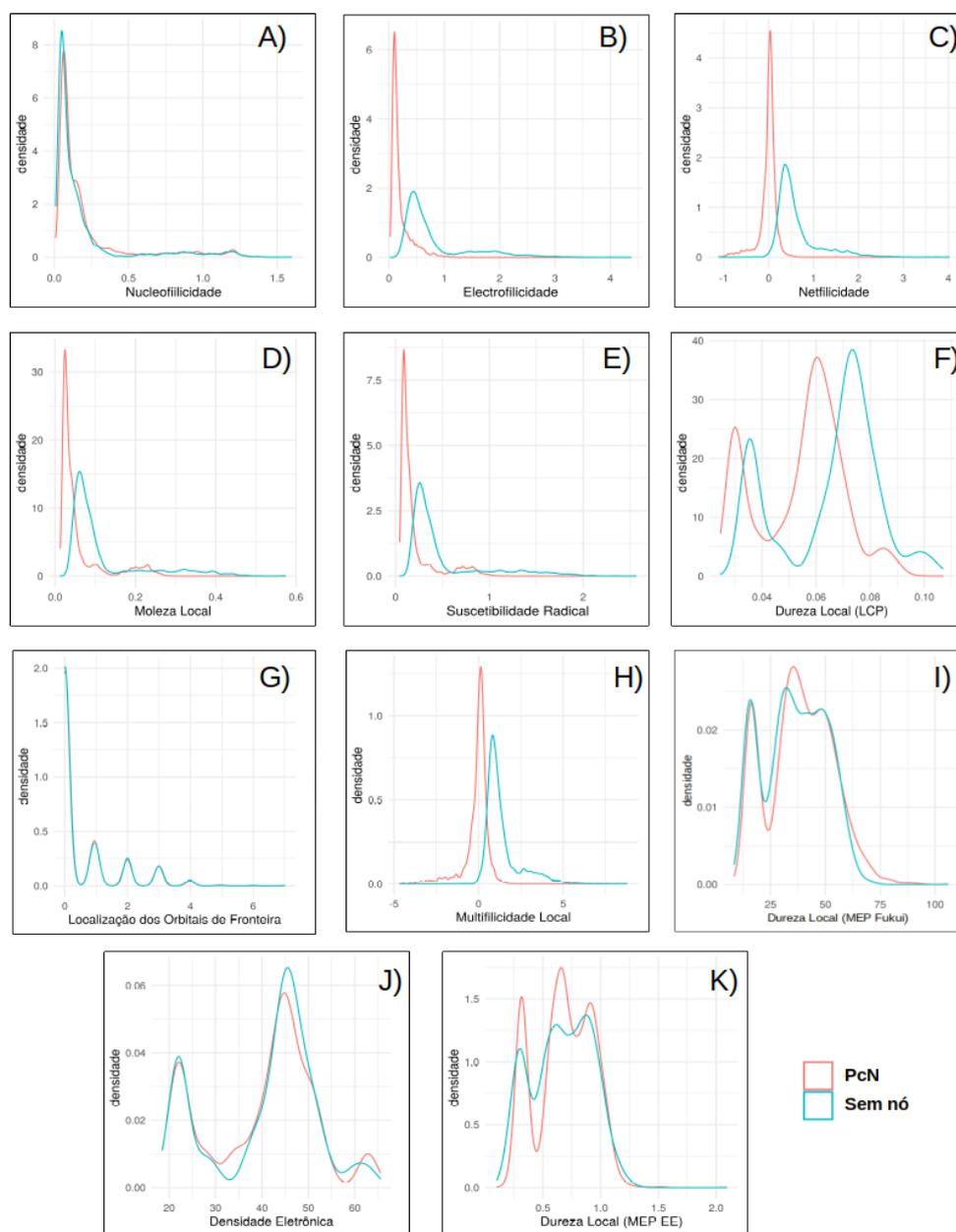


Figura 26 – Distribuição de densidade de probabilidade para os descritores de reatividade calculados entre as estruturas amostradas para 3KZN (PcN) e 4JQO (sem nó). A) Nucleofilicidade, B) Eletrofilicidade, C) Netfilicidade, D) Moleza Local, E) Suscetibilidade radical, F) Dureza Local (aproximação potencial químico local), G) Localização do Orbital Molecular, H) Multifilicidade, I) Dureza Local (aproximação potencial de Fukui), J) Densidade Eletrônica, K) Dureza Local (aproximação interação elétron-elétron).

Para os descritores que descrevem as interações controladas por carga, os métodos de dureza local, aquele baseado em potencial químico local (LCP) apresentou um deslocamento positivo em seus valores, com a diferença de padrão mais visível entre os dois tipos.

O mesmo tipo de análise buscando agrupar os dados pela temperatura da amostragem de dinâmica molecular, não apresentou nenhum padrão nos descritores calculados. Diante disso, como ainda existia várias fontes de possíveis efeitos, tais como diferentes resíduos, coordenadas tridimensionais, descritores quânticos, foi realizada uma análise exploratória com a Análise de Componentes Principais (PCA).

A PCA foi aplicada em duas rodadas: primeiro com todos os resíduos idênticos para as duas proteínas na região do nó; depois os resíduos de aminoácidos redundantes, isto é, quando o mesmo tipo aparecia mais de uma vez foram sendo retirados até que a maior parte da variância explicada começasse a diminuir significativamente. Na Figura 27 :A) e B) são apresentados os dados para a PCA inicial, agrupados pelo tipo de proteína e para o resíduo de aminoácido respectivamente, a variância explicada por componente principal (PC) em C) e os *loadings* para os primeiros cinco PCs em D).

Nas duas rodadas foram obtidas pelo menos 70% da variância explicada nos dois primeiros PCs. Com o PC1 com cerca de 50% e o PC2 com 22%. Para a segunda rodada de PCA, foram removidos os dados associados aos aminoácidos que apresentam baixa dispersão, principalmente horizontal, que é o eixo mais importante.

Na Figura 27: A), são apresentados os dados agrupados por tipo de proteína, mostrando que se trata de uma clara diferenciação dos descritores químicos quânticos obtidos a partir do tipo de proteína com nó e sem nó. Na Figura 27: B), foram agrupados os objetos em relação ao tipo de resíduo, mostrando também a separação. É possível afirmar que o padrão encontrado para o tipo de proteína é mais forte do que para o tipo de resíduo, pois seu deslocamento é no eixo horizontal. Além disso, se pode observar que todas as variáveis apresentam o mesmo sinal de carregamento para o primeiro PC, indicando que em geral os resíduos de aminoácidos da proteína sem nó são mais reativos, tanto nos descritores do tipo duro-duro quanto mole-mole.

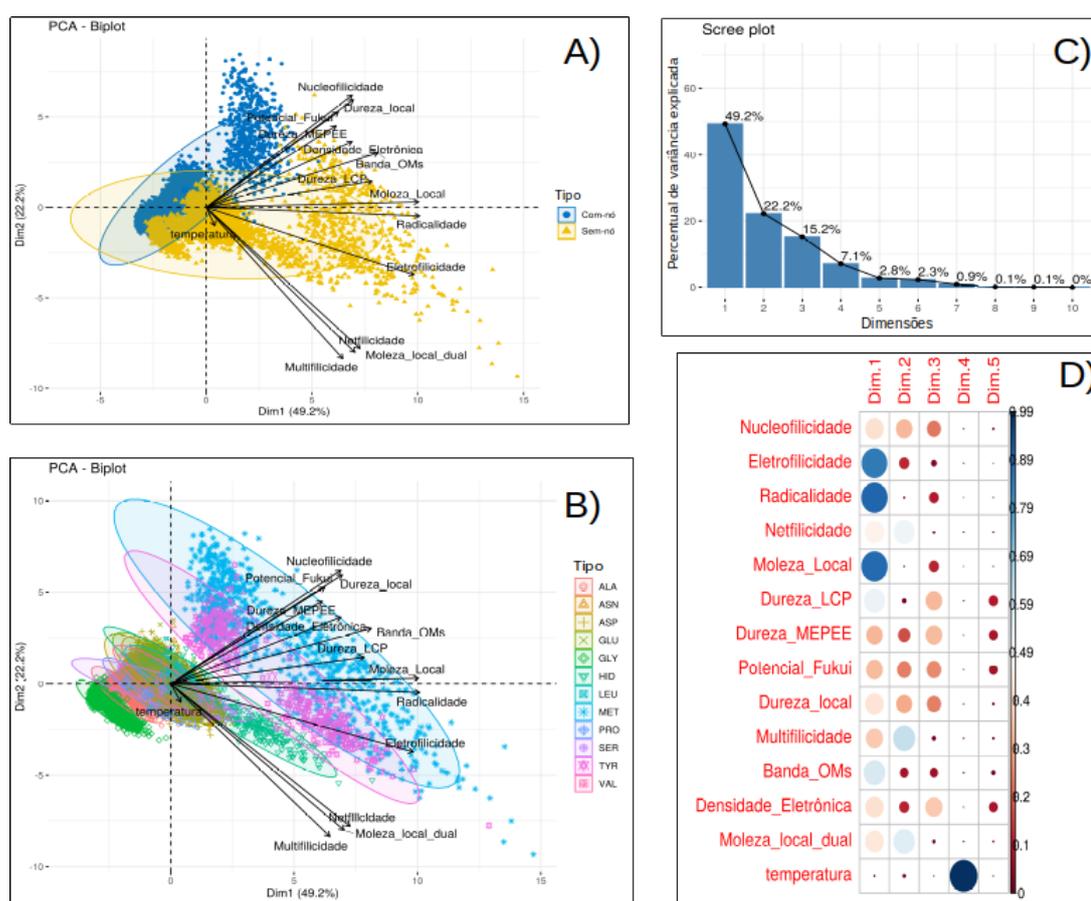


Figura 27 – Resultados da análise de componentes principais (PCA) empregando os dados obtidos pelos descritores de reatividade. A) Dados apresentados em relação ao tipo de proteína; B) Dados em relação ao tipo de resíduo de aminoácido; C) Variância explicada por componente principal; D) *Loadings* para os cinco primeiros componentes principais.

No segundo PC é onde ocorre a diferenciação entre os tipos de descritores, sendo Nucleofilicidade e Dureza_D local os maiores valores positivos para este PC e Multifilicidade, Moleza_Dual e Netfilicidade os valores mais negativos.

Esses descritores estão altamente associados ao PC que separa os dados por tipo de resíduo, mas com contribuição significativa também no primeiro PC. Por outro lado, a Moleza local e a suscetibilidade radical (abreviada como Radicalidade na Figura 27: B)) mostram as cargas mais altas em PC1 e quase zero em PC2, o que significa que podem representar a diferenciação dos dois tipos de proteínas praticamente sem relação do tipo resíduo. Também podemos observar que a temperatura utilizada na amostragem MD apresentou *loadings* próximos de zero nos dois PCs, sendo desprezíveis para as propriedades da estrutura eletrônica calculadas.

Na Figura 28, são apresentados os dados resultante para a segunda PCA, onde foram descartados os dados dos descritores da metade dos resíduos. Agora é possível ver mais claramente a separação dos dados por tipo de resíduos de aminoácidos. Além disso, com os dados menos redundantes foi possível identificar os conjuntos de dados onde ocorre a separação entre os tipos de proteínas com e sem nó, especificamente nos resíduos de Metionina, Tirosina e Histidina.

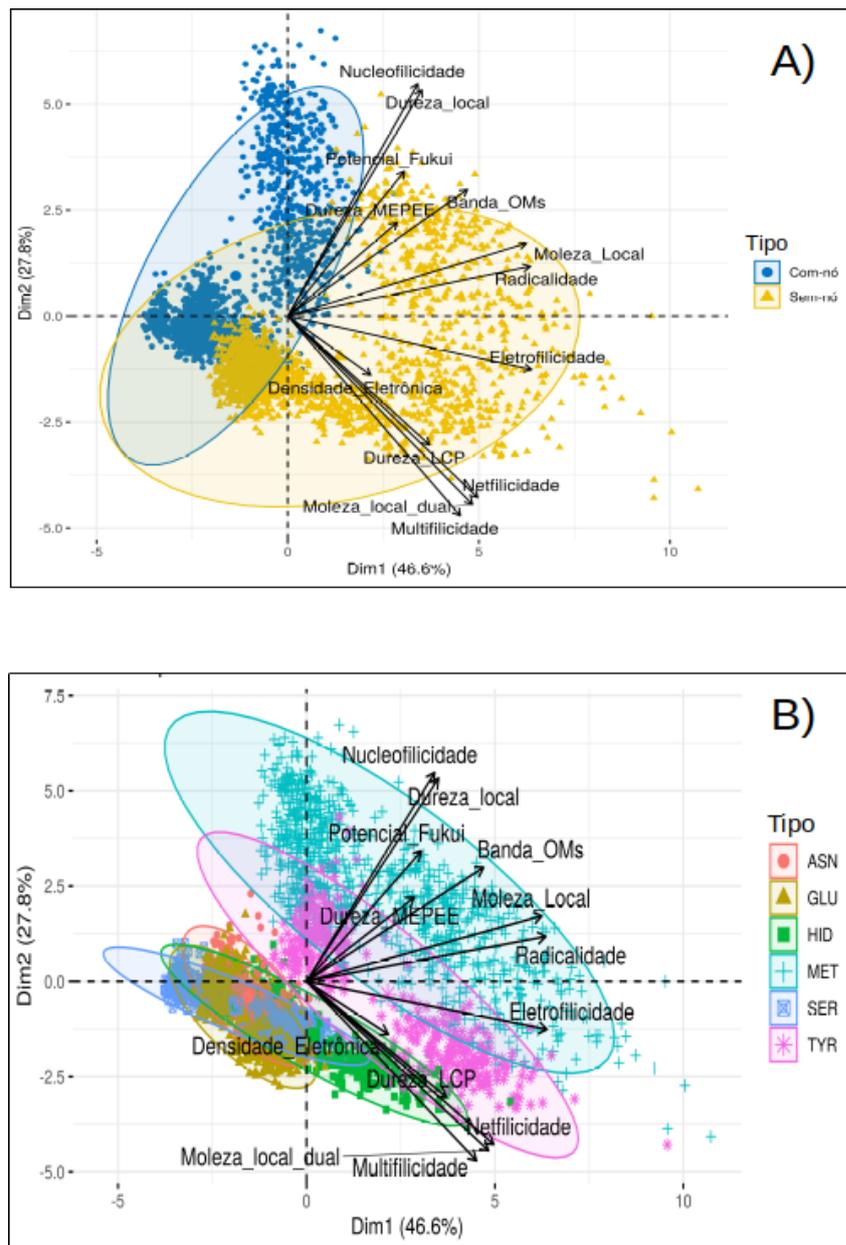


Figura 28 – Resultados da análise de componentes principais (PCA) para os descritores de reatividade. A) Objetos pelo tipo de proteína. B) Objetos pelo tipo de resíduo de aminoácido do nó.

As vantagens da análise de PCA é a possibilidade de explorar uma enorme matriz

de dados e dividir a variância por blocos de variáveis e classes de dados. A partir dessas análises até aqui, é possível afirmar que essas duas proteínas apresentam diferentes propriedades de estrutura eletrônica em seus resíduos conservados no nó.

Para complementar esta análise, foram construídos os mapas de calor da Figura 29 por meio dos valores médios de cada resíduo para todos os descritores. Em A para a PcN, e B a proteína sem nó. Os dados de entrada para os gráficos foram dimensionados pelo seu desvio padrão e centrados pela sua média, para que os descritores fossem comparáveis. Nestas figuras é possível visualizar o agrupamento entre os resíduos e descritores, e também como os valores médios dos descritores mudaram para alguns dos resíduos.

Por meio dos mapas de calor com os descritores agrupados por seu tipo, interações duro-duro e mole-mole, é possível verificar como não há um padrão definido na proteína com nó. Para a proteína sem nó, se observa um aumento na média dos resíduos de metionina, tirosina e histidina nos descritores de multifilicidade, netfilicidade e moleza. Além disso, os resíduos de metionina, que são dois no núcleo do nó, apresentam maiores valores do radical e da localização da banda OM para a proteína sem nó. Para o resíduo de tirosina na região analisada, os descritores de dureza locais são os mais altos nos dois tipos de estrutura.

Diante dessas observações, foi possível levantar algumas novas questões de pesquisa, como por exemplo, se esses arranjos geométricos dificultam a localização de orbitais moleculares de fronteira em seus átomos pertencentes e a mitigação de seu poder de eletrofilicidade/nucleofilicidade por meio de restrição de movimento específico. Além disso, se o nó geométrico pode tornar seus resíduos do núcleo menos reativos, isso pode refletir em um sítio ativo enzimático mais reativo e específico.

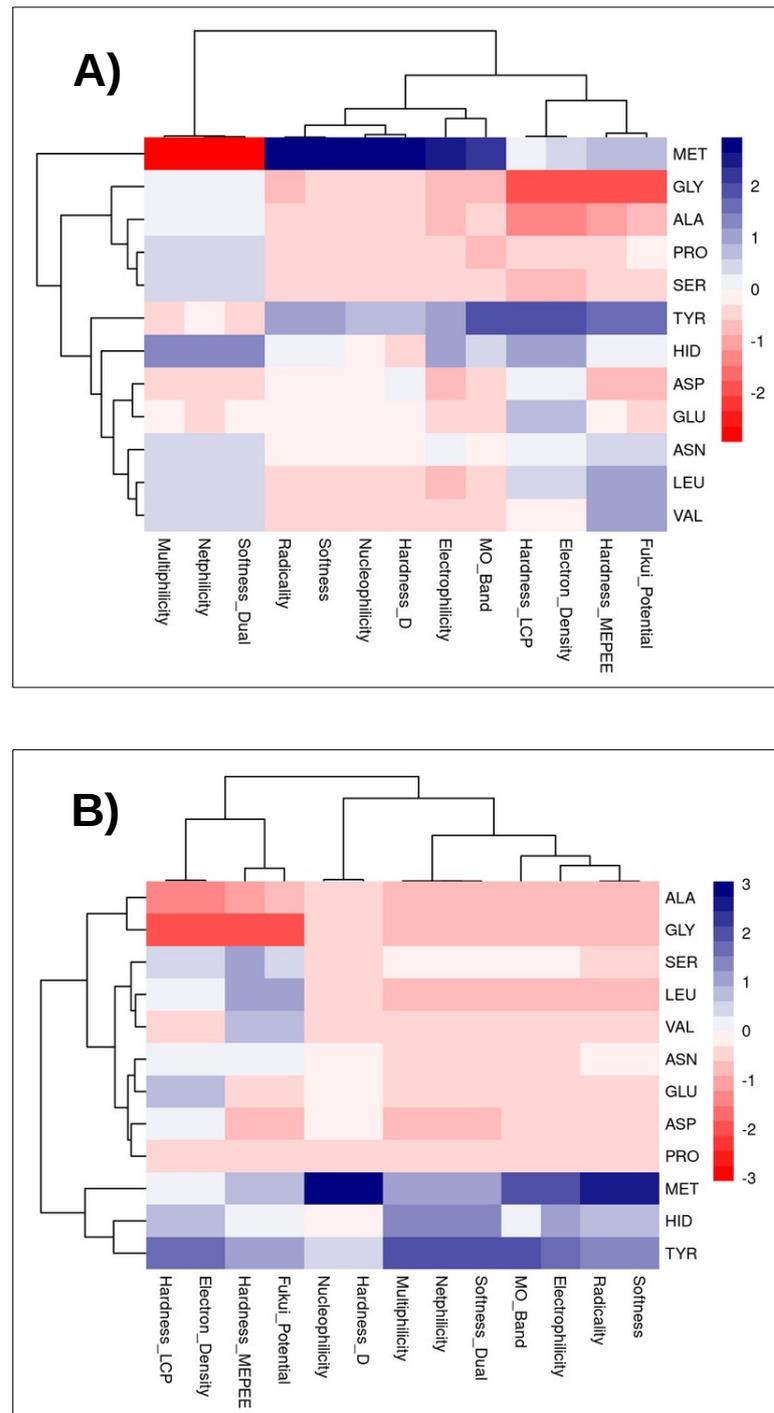


Figura 29 – Mapas de calor mostrando as médias dos descritores para cada resíduo analisado A) PcN B) Proteína sem nó.

A partir da PCA e dos mapas de calor foram selecionados os principais descritores para análise, além de revelar que o tipo de resíduo pode desempenhar um papel importante, mas secundário na descrição da variância dos dados. Assim, foram sele-

cionais quatro descritores para analisar a distribuição de cada resíduo comparando pelos dois tipos de proteína. Os resultados de tal análise são mostrados na Figura 30.

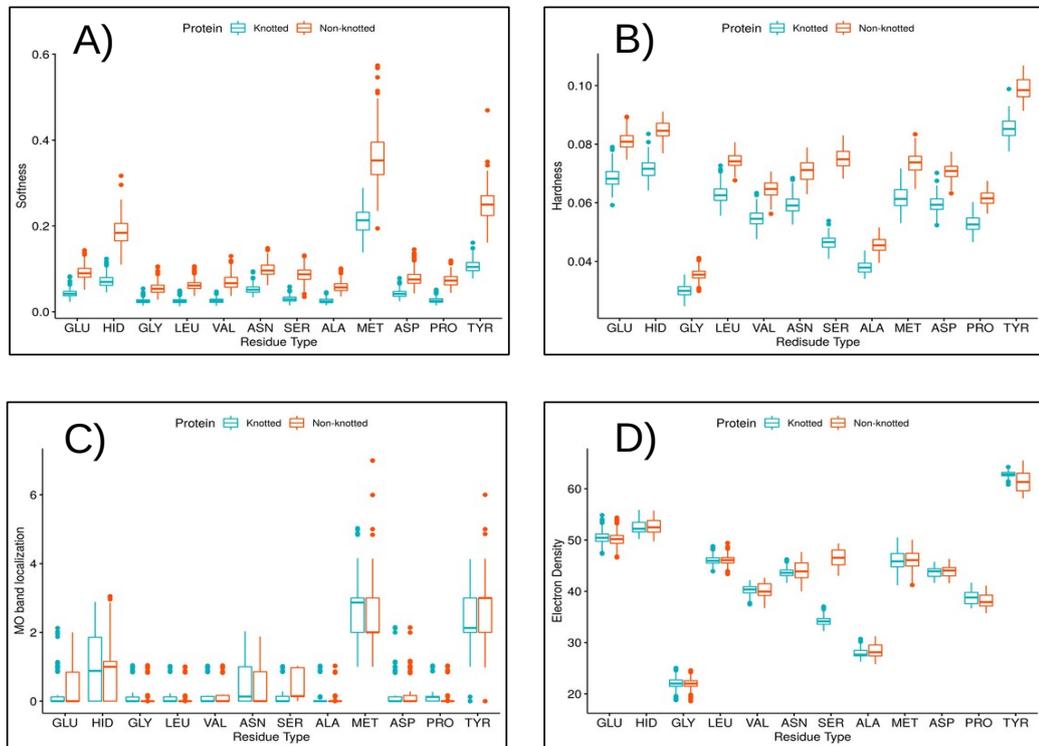


Figura 30 – Box-plot de alguns descritores químicos quânticos por resíduos de aminoácidos conservados no nó comparando entre os tipos de proteínas A) Moleza Local B) Dureza Local C) Localização de orbitais moleculares de fronteira D) Densidade eletrônica.

Para o descritor de moleza local, foi possível observar pela Figura 30: A) que os resíduos do tipo metionina, tirosina e histidina apresentaram as maiores diferenças em sua distribuição ao comparar as duas proteínas. O resíduo tirosina também aparece como tendo a maior média de dureza local (LCP), mas não sendo o único tipo de resíduo que apresentou distribuição diferente entre os tipos de proteínas, com o resíduo serina não apresentando nenhuma sobreposição de distribuição.

Para realizar uma análise estrutural, foi amostrado o conjunto mais representativo de coordenadas tridimensionais das simulações de dinâmica molecular a 310K, para visualizar espacialmente os descritores no núcleo do nó das duas proteínas. As equações de trabalho de dureza local são os descritores que mais graficamente mostram diferenças entre as duas proteínas, principalmente para a aproximação do Potencial

de Fukui. São apresentadas essas representações gráficas para a proteína com e sem nós nas Figuras 31 e 32.

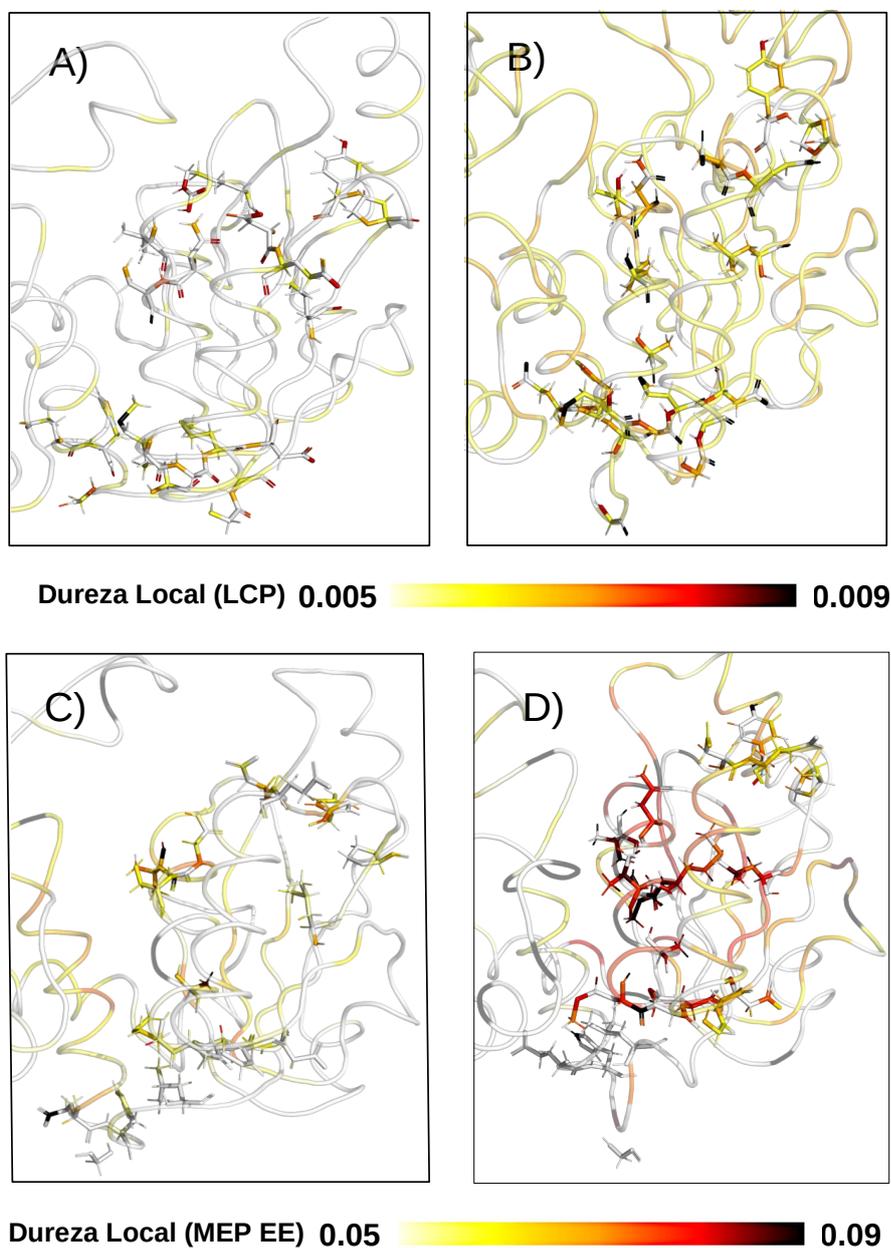


Figura 31 – Dureza local por átomos na região com nó em uma estrutura representativa: A) Dureza local com aproximação do potencial químico local para 3KZN (com nó) B) Dureza local com aproximação do potencial químico local para 4JQO (sem nó) C) Dureza local com aproximação interação elétron-elétron 3KZN (com nó) D) Dureza local com aproximação interação elétron-elétron 4JQO (sem nó).

Na Figura 31, são apresentados a dureza local para os átomos nas regiões de nó,

para a equação de trabalho usando as aproximações LCP e MEP EE, para a proteína com nó (Figura 31: A) e C)) e sem nó (Figura 31 :B) e D)). A equação de trabalho do LCP é muito dependente da densidade eletrônica e, portanto, os maiores valores tendem a estar nos átomos com mais elétrons. Assim, além de alguns altos valores de dureza local nos átomos dos resíduos da proteína com nó, a proteína sem nó apresenta os maiores valores em todos os resíduos da região equivalente ao nó. Além disso, a proteína sem nó apresenta maiores valores de dureza local calculados com as aproximações da interação elétron-elétron do que para a proteína com nó, conforme mostrado na Figura 31: C) e D).

Na Figura 32, são apresentadas duas representações da dureza local para os átomos na região de nó, para a equação do trabalho usando a aproximação do potencial de Fukui.

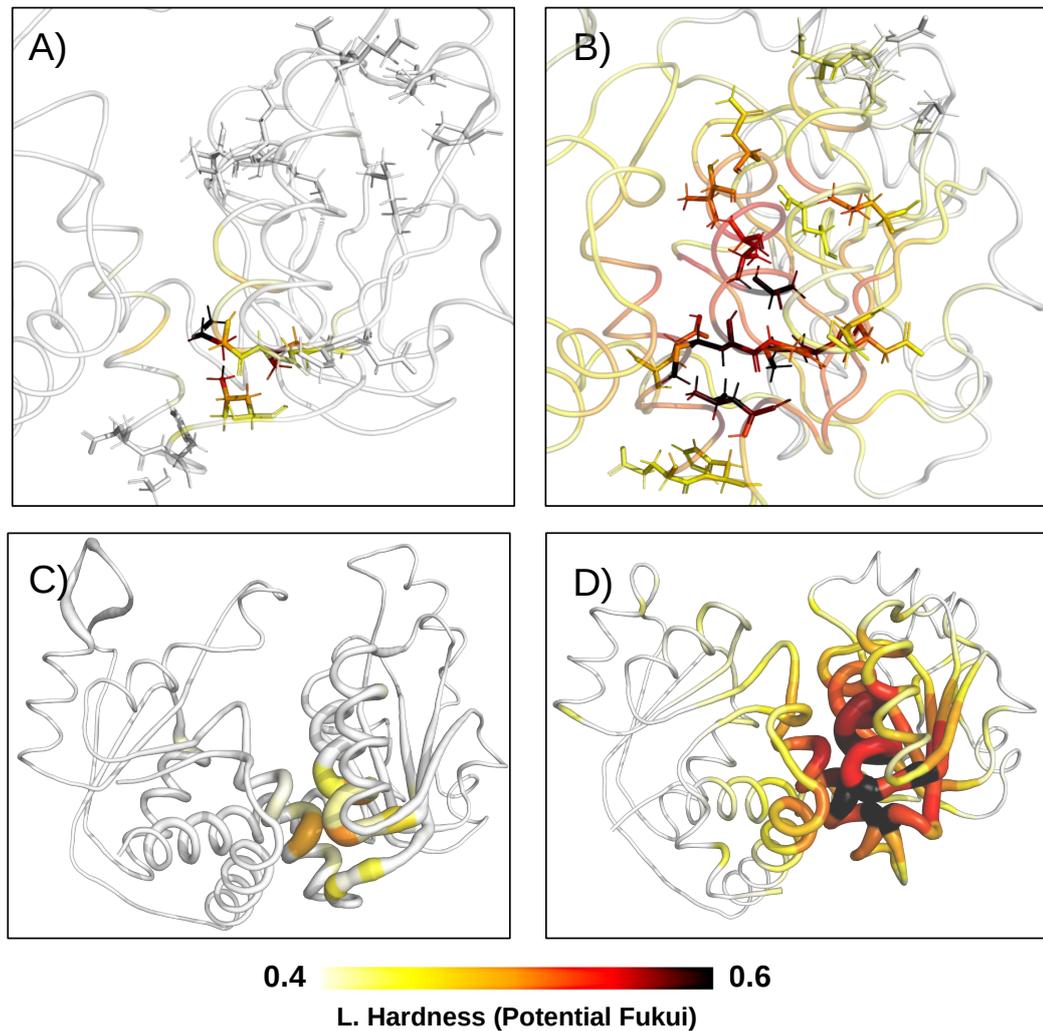


Figura 32 – Dureza local por átomos, obtida a partir do potencial de Fukui. Resíduos em destaque para a região do nó em A) 3KZN (PcN) e B) para 4JQO. Em C) 3KZN e D) 4JQO são mostradas as representações considerando toda a estrutura das proteínas.

A primeira sendo a fita e os palitos para os resíduos no núcleo do nó (Figura 32:A e Figura 32:B) e a segunda mostrando a representação em desenho da cadeia principal com a "espessura" dimensionada pelos valores do descritor. Foi possível observar que os maiores valores para este descritor estão nos resíduos da região do nó, com a proteína sem nó apresentando esses valores máximos em vários resíduos.

Com esses resultados, foram demonstrados que as propriedades eletrônicas e de reatividade calculadas para os resíduos do núcleo do nó são significativamente diferentes para cada uma das proteínas, embora sejam tão semelhantes tanto na estrutura terciária quanto na função biológica.

Embora não seja possível neste momento estabelecer uma relação direta entre essas diferenças e a presença de nós geométricos em proteínas em geral, as diferenças expostas por meio da investigação dos descritores de reatividade abre várias questões de pesquisa sobre como arranjos tridimensionais específicos dos resíduos de proteínas podem levar a mudanças em suas propriedades de reatividade.

Capítulo 6

Conclusão

Neste estudo foram usadas diferentes técnicas de simulação computacional para avaliar diferenças entre duas proteínas com alta similaridade no arranjo tridimensional, mas que diferem pela presença de um nó geométrico.

Com os dados obtidos por dinâmica molecular, o RMSF mostrou que, em todas as temperaturas avaliadas, as duas proteínas tiveram suas porções mais flexíveis fora da região do nó. As maiores flutuações foram observadas para a proteína sem nó (PDB 4JQO), que apresentou uma flutuação em maior destaque para o resíduo 243 (GLY243), dentro da região equivalente à região do nó na proteína com nó (PcN) (PDB 3KZN). A flutuação foi acentuada com elevação da temperatura para 373K e 436K para as duas estruturas. A fração de contatos nativos mostrou uma diminuição significativa apenas em 436K para as duas proteínas.

As análises de PCA mostraram que a proteína sem nó teve acesso a um conjunto mais variado de conformações, que pode apresentar menor resistência às mudanças conformacionais em comparação com a PcN. Os dados do DCCM mostraram que a proteína sem nó sofreu grandes mudanças em seus movimentos (correlação e anticorrelação) ao longo das mudanças de temperatura, o que sugere que ela está sujeita a maiores variações conformacionais, indicando menor estabilidade.

Em relação ao estudo utilizando descritores de reatividade (DR) foi observado que a proteína sem nós mostrou-se propensa a maior reatividade em relação a descritores usados que descrevem interações do tipo mole-mole ou duro-duro.

Referências

- 1 DABROWSKI-TUMANSKI, P.; STASIAK, A.; SULKOWSKA, J. I. In search of functional advantages of knots in proteins. *PLoS one*, Public Library of Science, v. 11, n. 11, 2016.
- 2 SUŁKOWSKA, J. I.; NOEL, J. K.; ONUCHIC, J. N. Energy landscape of knotted protein folding. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 109, n. 44, p. 17783–17788, 2012.
- 3 XU, Y. et al. Stabilizing effect of inherent knots on proteins revealed by molecular dynamics simulations. *Biophysical journal*, Elsevier, v. 115, n. 9, p. 1681–1689, 2018.
- 4 HSU, S.-T. D. et al. Tying up the loose ends: a mathematically knotted protein. *Frontiers in chemistry*, Frontiers, v. 9, p. 328, 2021.
- 5 MARZZOCO, A.; TORRES, B. B. Bioquímica básica. p. 392–392, 2017.
- 6 LEHNINGER, A. L.; NELSON, D.; COX, M. Princípios de bioquímica, 4. edição. *Sarvier*, São Paulo, 2006.
- 7 JÚNIOR, W.; FRANCISCO, W. Proteínas: Hidrólise, precipitação e um tema para o ensino de química. *São Paulo*, 2006.
- 8 FAÍSCA, P. F. Knotted proteins: A tangled tale of structural biology. *Computational and structural biotechnology journal*, Elsevier, v. 13, p. 459–468, 2015.
- 9 LIM, N. C.; JACKSON, S. E. Molecular knots in biology and chemistry. *Journal of Physics: Condensed Matter*, IOP Publishing, v. 27, n. 35, p. 354101, 2015.
- 10 VERLI, H. Bioinformática: da biologia à flexibilidade molecular. Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.
- 11 BUSTAMANTE, A. et al. The energy cost of polypeptide knot formation and its folding consequences. *Nature communications*, Nature Publishing Group, v. 8, n. 1, p. 1–8, 2017.
- 12 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Vmd: visual molecular dynamics. *Journal of molecular graphics*, Elsevier, v. 14, n. 1, p. 33–38, 1996.
- 13 BECCARA, S. a et al. Folding pathways of a knotted protein with a realistic atomistic force field. *PLoS Comput Biol*, Public Library of Science, v. 9, n. 3, p. e1003002, 2013.

- 14 SRIRAMOJU, M. K.; YANG, T.-J.; HSU, S.-T. D. Comparative folding analyses of unknotted versus trefoil-knotted ornithine transcarbamylases suggest stabilizing effects of protein knots. *Biochemical and biophysical research communications*, Elsevier, v. 503, n. 2, p. 822–829, 2018.
- 15 XU, Y. et al. Revealing cooperation between knotted conformation and dimerization in protein stabilization by molecular dynamics simulations. *The journal of physical chemistry letters*, ACS Publications, v. 10, n. 19, p. 5815–5822, 2019.
- 16 DABROWSKI-TUMANSKI, P. et al. Knotprot 2.0: a database of proteins with knots and other entangled structures. *Nucleic acids research*, Oxford University Press, v. 47, n. D1, p. D367–D375, 2019.
- 17 ROBSON, B. De novo protein folding on computers. benefits and challenges. *Computers in Biology and Medicine*, Elsevier, p. 105292, 2022.
- 18 RICHARDSON, J. S. β -sheet topology and the relatedness of proteins. *Nature*, Nature Publishing Group, v. 268, n. 5620, p. 495–500, 1977.
- 19 MANSFIELD, M. L. Are there knots in proteins? *Nature structural biology*, Nature Publishing Group, v. 1, n. 4, p. 213–214, 1994.
- 20 PAISSONI, C. et al. Converging experimental and computational views of the knotting mechanism of a small knotted protein. *Biophysical Journal*, Elsevier, v. 120, n. 11, p. 2276–2286, 2021.
- 21 BERMAN, H. M. et al. The protein data bank. *Nucleic acids research*, Oxford University Press, v. 28, n. 1, p. 235–242, 2000.
- 22 JACKSON, S. E.; SUMA, A.; MICHELETTI, C. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current opinion in structural biology*, Elsevier, v. 42, p. 6–14, 2017.
- 23 DABROWSKI-TUMANSKI, P.; JARMOLINSKA, A.; SULKOWSKA, J. Prediction of the optimal set of contacts to fold the smallest knotted protein. *Journal of Physics: Condensed Matter*, IOP Publishing, v. 27, n. 35, p. 354109, 2015.
- 24 DZUBIELLA, J. Sequence-specific size, structure, and stability of tight protein knots. *Biophysical journal*, Elsevier, v. 96, n. 3, p. 831–839, 2009.
- 25 HORNER, K. E. et al. Knot theory in modern chemistry. *Chemical Society Reviews*, Royal Society of Chemistry, v. 45, n. 23, p. 6432–6448, 2016.
- 26 JAMROZ, M. et al. Knotprot: a database of proteins with knots and slipknots. *Nucleic acids research*, Oxford University Press, v. 43, n. D1, p. D306–D314, 2015.
- 27 ADAMS, C. C. *The knot book*. [S.l.]: American Mathematical Soc., 1994.
- 28 EXEL, R. Von neumann e a teoria de álgebra de operadores. *Estudos Avançados*, SciELO Brasil, v. 10, n. 26, p. 211–225, 1996.
- 29 KONIARIS, K.; MUTHUKUMAR, M. Self-entanglement in ring polymers. *The Journal of chemical physics*, American Institute of Physics, v. 95, n. 4, p. 2873–2881, 1991.

- 30 KOLESOV, G. et al. Protein knot server: detection of knots in protein structures. *Nucleic acids research*, Oxford University Press, v. 35, n. suppl_2, p. W425–W428, 2007.
- 31 ALEXANDER, K.; TAYLOR, A. J.; DENNIS, M. R. Proteins analysed as virtual knots. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–10, 2017.
- 32 PETTERSEN, E. F. et al. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, Wiley Online Library, v. 25, n. 13, p. 1605–1612, 2004.
- 33 VIRNAU, P.; MIRNY, L. A.; KARDAR, M. Intricate knots in proteins: Function and evolution. *PLoS Comput Biol*, Public Library of Science, v. 2, n. 9, p. e122, 2006.
- 34 SUŁKOWSKA, J. I. et al. Conservation of complex knotting and slipknotting patterns in proteins. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 109, n. 26, p. E1715–E1723, 2012.
- 35 SUŁKOWSKA, J. I. et al. Stabilizing effect of knots on proteins. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 105, n. 50, p. 19714–19719, 2008.
- 36 ZHAO, Y. et al. The exclusive effects of chaperonin on the behavior of proteins with 52 knot. *PLoS computational biology*, Public Library of Science, v. 14, n. 3, p. e1005970, 2018.
- 37 JARMOLINSKA, A. I. et al. Proteins' knotty problems. *Journal of molecular biology*, Elsevier, v. 431, n. 2, p. 244–257, 2019.
- 38 BÖLINGER, D. et al. A stevedore's protein knot. *PLoS computational biology*, Public Library of Science, v. 6, n. 4, p. e1000731, 2010.
- 39 WALLIN, S.; ZELDOVICH, K. B.; SHAKHNOVICH, E. I. The folding mechanics of a knotted protein. *Journal of molecular biology*, Elsevier, v. 368, n. 3, p. 884–893, 2007.
- 40 MORGON, N. H. *Métodos de química teórica e modelagem molecular*. [S.l.]: Editora Livraria da Física, 2007.
- 41 NAMBA, A. M.; SILVA, V. B. da; SILVA, C. D. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. *Eclética Química*, SciELO Brasil, v. 33, n. 4, p. 13–24, 2008.
- 42 PERLINSKA, A. P. et al. Mg²⁺-dependent methyl transfer by a knotted protein: A molecular dynamics simulation and quantum mechanics study. *ACS catalysis*, ACS Publications, v. 10, n. 15, p. 8058–8068, 2020.
- 43 GRILLO, I. B. et al. Elucidating enzymatic catalysis using fast quantum chemical descriptors. *Journal of Chemical Information and Modeling*, ACS Publications, v. 60, n. 2, p. 578–591, 2020.
- 44 ROCHA-SANTOS, A. et al. Thermochemical and quantum descriptor calculations for gaining insight into ricin toxin a (rta) inhibitors. *ACS omega*, ACS Publications, v. 6, n. 13, p. 8764–8777, 2021.

- 45 ROCHA, R. E. et al. A higher flexibility at the sars-cov-2 main protease active site compared to sars-cov and its potentialities for new inhibitor virtual screening targeting multi-conformers. *Journal of Biomolecular Structure and Dynamics*, Taylor & Francis, p. 1–21, 2021.
- 46 FAVER, J.; JR, K. M. M. Utility of the hard/soft acid- base principle via the fu-kui function in biological systems. *Journal of chemical theory and computation*, ACS Publications, v. 6, n. 2, p. 548–559, 2010.
- 47 GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; ROCHA, G. B. Primordia: A software to explore reactivity and electronic structure in large biomolecules. *Journal of Chemical Information and Modeling*, ACS Publications, v. 60, n. 12, p. 5885–5890, 2020.
- 48 GEERLINGS, P.; PROFT, F. D.; LANGENAEKER, W. Conceptual density functional theory. *Chemical reviews*, ACS Publications, v. 103, n. 5, p. 1793–1874, 2003.
- 49 PARR, R. G. et al. Electronegativity: the density functional viewpoint. *The Journal of Chemical Physics*, American Institute of Physics, v. 68, n. 8, p. 3801–3807, 1978.
- 50 PEARSON, R. G. Hard and soft acids and bases—the evolution of a chemical concept. *Coordination chemistry reviews*, Elsevier, v. 100, p. 403–425, 1990.
- 51 SUŁKOWSKA, J. I.; SUŁKOWSKI, P.; ONUCHIC, J. Dodging the crisis of folding proteins with knots. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 106, n. 9, p. 3119–3124, 2009.
- 52 NOEL, J. K.; ONUCHIC, J. N.; SUŁKOWSKA, J. I. Knotting a protein in explicit solvent. *The Journal of Physical Chemistry Letters*, ACS Publications, v. 4, n. 21, p. 3570–3573, 2013.
- 53 WU, X. et al. Folding mechanisms of trefoil knot proteins studied by molecular dynamics simulations and go-model. In: *Advance in Structural Bioinformatics*. [S.l.]: Springer, 2015. p. 93–110.
- 54 SRIRAMOJU, M. K. et al. Topologically knotted deubiquitinases exhibit unprecedented mechanostability to withstand the proteolysis by an aaa+ protease. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–9, 2018.
- 55 MARTÍN, Á. S. et al. Knots can impair protein degradation by atp-dependent proteases. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 114, n. 37, p. 9864–9869, 2017.
- 56 WOJCIECHOWSKI, M. et al. Unfolding knots by proteasome-like systems: Simulations of the behaviour of folded and neurotoxic proteins. *Molecular bioSystems*, Royal Society of Chemistry, v. 12, n. 9, p. 2700–2712, 2016.
- 57 FONSEKA, H. Y. Y. et al. Unfolding and translocation of knotted proteins by clp biological nanomachines: Synergistic contribution of primary sequence and topology revealed by molecular dynamics simulations. *The Journal of Physical Chemistry B*, ACS Publications, v. 125, n. 27, p. 7335–7350, 2021.
- 58 SIVERTSSON, E. M.; JACKSON, S. E.; ITZHAKI, L. S. The aaa+ protease clpxp can easily degrade a 31 and a 52-knotted protein. *Scientific reports*, Nature Publishing Group, v. 9, n. 1, p. 1–14, 2019.

- 59 SRIRAMOJU, M. K.; CHEN, Y.; HSU, S.-T. D. Protein knots provide mechano-resilience to an aaa+ protease-mediated proteolysis with profound atp energy expenses. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, Elsevier, v. 1868, n. 2, p. 140330, 2020.
- 60 SHI, D. et al. Crystal structure of human ornithine transcarbamylase complexed with carbamoyl phosphate and l-norvaline at 1.9 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 39, n. 4, p. 271–277, 2000.
- 61 SHI, D. et al. Structures of n-acetylornithine transcarbamoylase from xanthomonas campestris complexed with substrates and substrate analogs imply mechanisms for substrate binding and catalysis. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 64, n. 2, p. 532–542, 2006.
- 62 PHILLIPS, J. C. et al. Scalable molecular dynamics with namd. *Journal of computational chemistry*, Wiley Online Library, v. 26, n. 16, p. 1781–1802, 2005.
- 63 CASE, D. et al. Amber 2018; 2018. *University of California, San Francisco*, 2018.
- 64 DOLINSKY, T. J. et al. Pdb2pqr: an automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic acids research*, Oxford University Press, v. 32, n. suppl_2, p. W665–W667, 2004.
- 65 MAIER, J. A. et al. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, ACS Publications, v. 11, n. 8, p. 3696–3713, 2015.
- 66 RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, Elsevier, v. 23, n. 3, p. 327–341, 1977.
- 67 FELLER, S. E. et al. Constant pressure molecular dynamics simulation: the langevin piston method. *The Journal of chemical physics*, American Institute of Physics, v. 103, n. 11, p. 4613–4621, 1995.
- 68 CONTESSOTO, V. d. G. et al. Introdução ao problema de enovelamento de proteínas: uma abordagem utilizando modelos computacionais simplificados. *Revista Brasileira de Ensino de Física*, SciELO Brasil, v. 40, n. 4, 2018.
- 69 MCGIBBON, R. T. et al. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical journal*, Elsevier, v. 109, n. 8, p. 1528–1532, 2015.
- 70 GRANT, B. J. et al. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, v. 22, n. 21, p. 2695–2696, 08 2006. ISSN 1367-4803. Disponível em: <<https://doi.org/10.1093/bioinformatics/btl461>>.
- 71 DAVID, C. C.; JACOBS, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. In: *Protein dynamics*. [S.l.]: Springer, 2014. p. 193–226.

- 72 WEI-YA, L. et al. Exploring the cause of the inhibitor 4ax attaching to binding site disrupting protein tyrosine phosphatase 4a1 trimerization by molecular dynamic simulation. *Journal of Biomolecular Structure and Dynamics*, Taylor & Francis, v. 37, n. 18, p. 4840–4851, 2019.
- 73 SKJÆRVEN, L. et al. Integrating protein structural dynamics and evolutionary analysis with bio3d. *BMC bioinformatics*, BioMed Central, v. 15, n. 1, p. 1–11, 2014.
- 74 NDAGI, U.; MHLONGO, N. N.; SOLIMAN, M. E. The impact of thr91 mutation on c-src resistance to um-164: molecular dynamics study revealed a new opportunity for drug design. *Molecular Biosystems*, Royal Society of Chemistry, v. 13, n. 6, p. 1157–1171, 2017.
- 75 ZHOU, L. et al. Identification of the potential dual inhibitor of protein tyrosine phosphatase sigma and leukocyte common antigen-related phosphatase by virtual screen, molecular dynamic simulations and post-analysis. *Journal of Biomolecular Structure and Dynamics*, Taylor & Francis, v. 39, n. 1, p. 45–62, 2021.
- 76 WANG, R.-R. et al. Exploring the reason for increased activity of shp2 caused by d61y mutation through molecular dynamics. *Computational biology and chemistry*, Elsevier, v. 78, p. 133–143, 2019.
- 77 STEWART, J. J. Mopac: a semiempirical molecular orbital program. *Journal of computer-aided molecular design*, Springer, v. 4, n. 1, p. 1–103, 1990.
- 78 STEWART, J. J. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *Journal of molecular modeling*, Springer, v. 19, n. 1, p. 1–32, 2013.
- 79 STEWART, J. J. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *International Journal of Quantum Chemistry*, Wiley Online Library, v. 58, n. 2, p. 133–146, 1996.
- 80 STEIN, S. A. M. et al. Principal components analysis: a review of its application on molecular dynamics data. *Annual Reports in Computational Chemistry*, Elsevier, v. 2, p. 233–261, 2006.
- 81 SUN, Y.-Z. et al. Exploring the effect of aplidin on low molecular weight protein tyrosine phosphatase by molecular docking and molecular dynamic simulation study. *Computational Biology and Chemistry*, Elsevier, v. 83, p. 107123, 2019.
- 82 SUN, Y.-Z. et al. Exploring the effect of n308d mutation on protein tyrosine phosphatase-2 cause gain-of-function activity by a molecular dynamics study. *Journal of Cellular Biochemistry*, Wiley Online Library, v. 120, n. 4, p. 5949–5961, 2019.

Apêndice **A**

Apêndice

- Artigo publicado na revista *Journal of Molecular Modeling*.



Investigation of the structural dynamics of a knotted protein and its unknotted analog using molecular dynamics

José Cícero Alves Silva¹ · Elton José Ferreira Chaves² · Gabriel Aires Urquiza de Carvalho³ · Gerd Bruno Rocha¹

Received: 22 December 2021 / Accepted: 12 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The role of knots in proteins remains elusive. Some studies suggest an impact on stability; the difficulty in comparing systems to assess this effect, however, has been a significant challenge. In this study, we produced and analyzed molecular dynamic trajectories considering three different temperatures of two variants of ornithine transcarbamylase (OTC), only one of which has a 3_1 knot, in order to evaluate the relative stability of the two molecules. RMSD showed equilibrated structures for the produced trajectories, and RMSF showed subtle differences in flexibility. In the knot moiety, the knotted protein did not show a great deal of fluctuation at any temperature. For the unknotted protein, the residue GLY243 showed a high fluctuation in the corresponding moiety. The fraction of native contacts (Q) showed a similar profile at all temperatures, with the greatest decrease by 436 K. The investigation of conformational behavior with principal component analysis (PCA) and dynamic cross-correlation map (DCCM) showed that knotted protein is less likely to undergo changes in its conformation under the conditions employed compared to unknotted. PCA data showed that the unknotted protein had greater dispersion in its conformations, which suggests that it has a greater capacity for conformation transitions in response to thermal changes. DCCM graphs comparing the 310 K and 436 K temperatures showed that the knotted protein had less change in its correlation and anti-correlation movements, indicating stability compared to the unknotted.

Keywords Knotted proteins · Stability · Molecular dynamics · PCA · DCCM

Introduction

Knotted proteins represent an intriguing topic in biochemistry. Evidence concerning the existence of knots in the structure of proteins remains challenging even 40 years after their existence was first reported [1, 2]. There is great speculation about the role of the knot in proteins, whether this is an advantage or disadvantage for their correct functioning [3–6]. Since the first systematic study carried out by Mansfield [2] in 1994, few proteins with this structural characteristic have been verified. Currently, in the Protein

Data Bank (PDB) [7], less than 2% of structures have at least one structural knot [6, 8]. Recent studies have pointed to a slower folding kinetics for knotted proteins [9–11]. Another aspect of relevance is whether a knot provides greater stability to the protein or influences its biological activity [9, 12–15]. This topic is very open to research and study for the establishment of the real importance of the knot in proteins.

Another important point is the mechanical resistance of knotted proteins to the controlled degradation of non-functional proteins [16–18], since this is undoubtedly one of the crucial processes in cells. Recently, in a study with different topologies of knots in proteins, a higher resistance to degradation was observed in comparison to unknotted variants [19]. Different types of knots have different effects in relation to degradation, which can occur in a partial way or formation of intermediates [20]. In this step, the knot has an influence when in contact with the channel responsible for degradation, as observed in the protease AAA+ ClpXP [17], and a mechano-stability effect has already been observed for knotted proteins [16, 21]. This reaffirms the interest in

✉ Gerd Bruno Rocha
gbr@quimica.ufpb.br

¹ Department of Chemistry, Federal University of Paraíba, Cidade Universitária, Joao Pessoa, PB 58051-900, Brazil

² Department of Biotechnology, Federal University of Paraíba, Cidade Universitária, Joao Pessoa, PB 58051-900, Brazil

³ Fundamental Department of Chemistry, Federal University of Pernambuco, Cidade Universitária, Recife, PE 50670-901, Brazil

establishing the impact that the presence of the knot brings to the proteins.

For studying knotted proteins, it is very important to identify the knot present in the structure. Currently, knotted proteins are identified with knot types 3_1 , 4_1 , 5_2 , and 6_1 [13, 22–24]. Among all the identified knotted proteins, those with knot 3_1 (known as the trefoil knot [25]) are the most numerous [12, 26]. The 3_1 knot is found in structures of proteins with different functions [27], such as carbonic anhydrase (PDB 3MDZ: A), RNA methyltransferase (PDB 4H3Z: B), and carbamoyltransferase (PDB 3KZK: A). The knot is identified using online servers that have tools to locate and identify the type of knot present in a knotted protein, for example, KnotProt [28].

Most of the research has focused on studying knotted proteins with knot 3_1 , which is formed by only three crosses [29–32]. The MJ0366 protein with 92 amino acids has also been systematically studied as it is the smallest known protein that contains a knot, which facilitates computational simulation and information extraction [29, 33, 34]. For some of these, simplified structure models have been used.

An important question is this: does the knot have any relevance to protein stability? To address this question, the work carried out by Xu and collaborators [6] sought to contribute by promoting the computational removal of the knot from the structure of a knotted protein. A further meaningful computational experiment might be the comparison between similar structures that occur in nature, rather than using a computationally manipulated structure, comparing it to an unedited native conformation. The ornithine transcarbamylase (OTC) family contains important proteins that act in the biosynthesis of arginine in bacteria and plants and the urea

cycle in mammals [35]. Among them, we find a perfect test case to carry out this comparison, PDB entries 3KZN and 4JQO, which correspond to a knotted and unknotted protein, respectively. Previous study, using only the PDB structures for 3KZN and 4JQO, did not show significant differences between the two proteins [4].

In the current study, we performed molecular dynamic simulations to investigate structural characteristics for PDB 3KZN and PDB 4JQO, considering three different temperatures. We employ principal component analysis and cross-correlation maps. We found greater resistance to change in the conformations for knotted protein.

Materials and methods

In Fig. 1, we present a schematic overview of the followed methodology in this study.

Knotted and unknotted protein structures

The structures of the two OTC variants, shown in Fig. 2, were obtained from the Protein Data Bank [7] and represent entries PDB 3KZN and PDB 4JQO, corresponding to a knotted and an unknotted protein, respectively. Figure 2 indicates that the two proteins present very similar tertiary structures. In Fig. 2a, the PDB 3KZN is shown, and in Fig. 2b, the structure for the PDB 4JQO. Highlighted in blue, each protein has two similarly spaced β -sheet moieties. Just below the three-dimensional structures, the annotated primary sequence is shown (more details are available in the supplementary information of this paper). The three-dimensional

Fig. 1 Workflow of the methods employed in this study

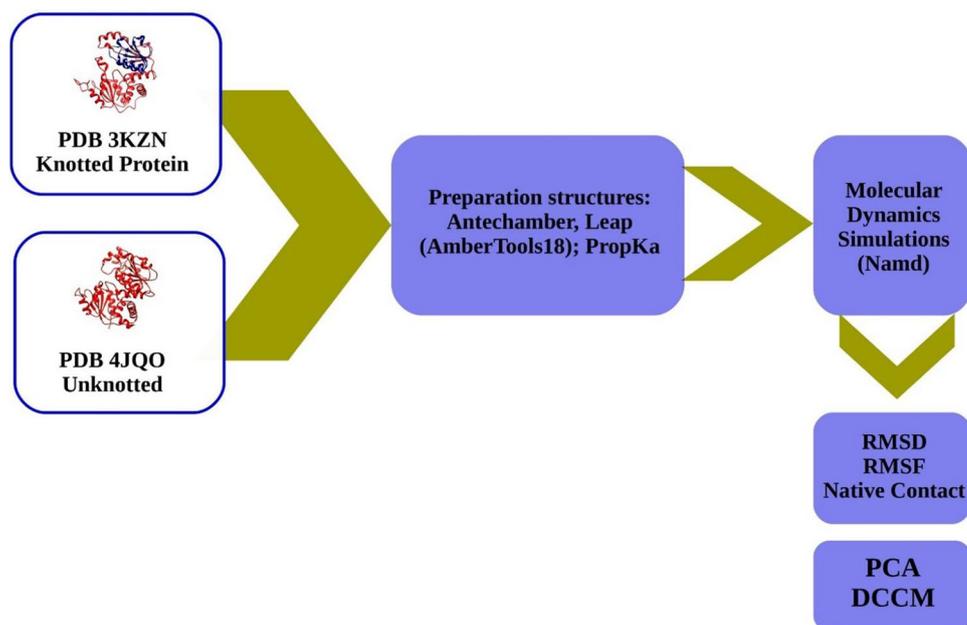
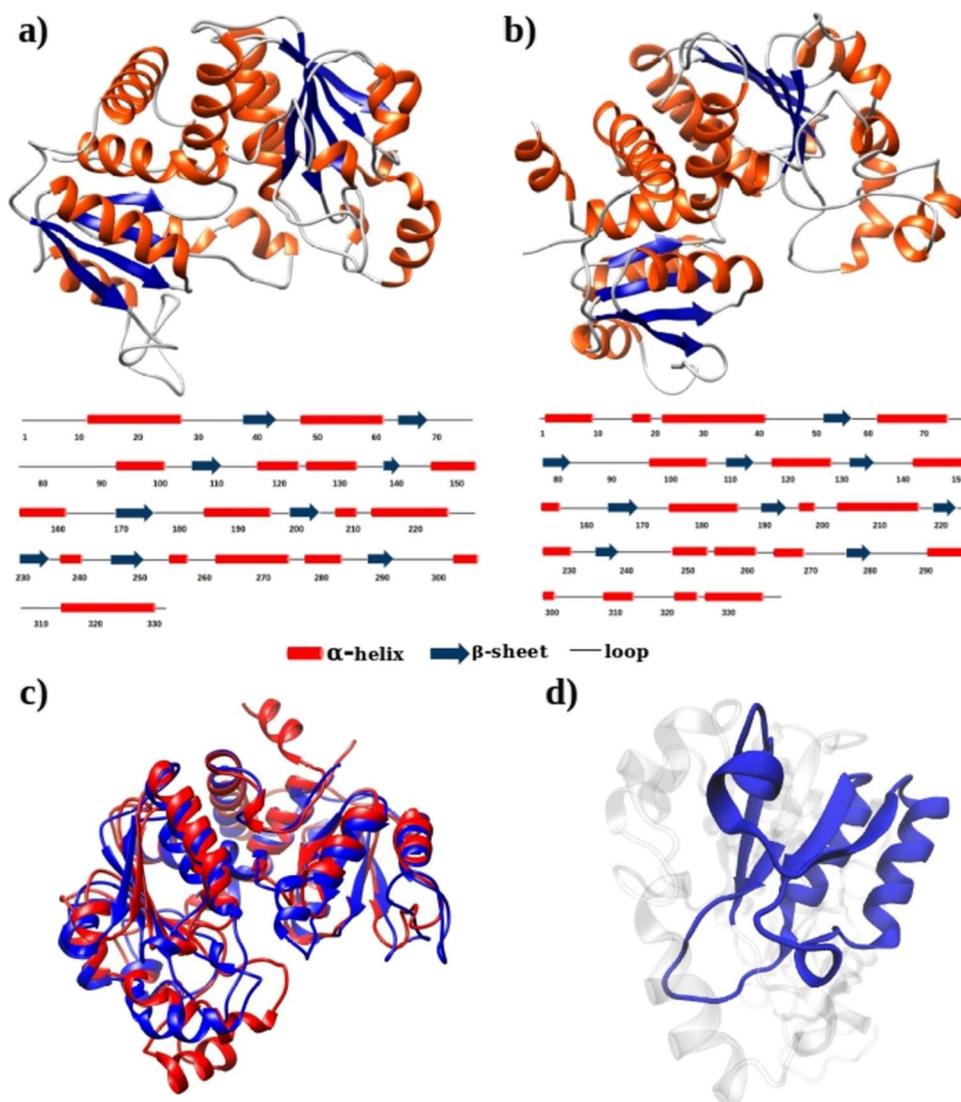


Fig. 2 Protein structures: **a** PDB 3KZN (knotted protein) and **b** PDB 4JQO (unknotted). **c** The alignment of the structures and **d** knot moiety is shown



superposition of the proteins is shown in Fig. 2c. The knotted protein is shown in blue and the unknotted in red. In Fig. 2d, the knot moiety is highlighted in blue for PDB 3KZN. It is clear from these figures that the structural similarity between both proteins is indeed great.

Molecular dynamic simulations

The two proteins PDB 3KZN (knotted protein) and PDB 4JQO (unknotted) were subjected to molecular dynamic simulations using NAMD v2.12 software [37]. Initially, we used the Antechamber software in the AmberTools18 [38] package to parametrize the non-standard KCX residue, an acetylated lysine present in the PDB 3KZN. The PropKa server [39] was used to adjust the protonation states of the residues in order to simulate a protein at pH 7.4. Then, the Leap software from the AmberTools18 [38] package was used for further pre-treatment of the systems. We carried

out solvation for both proteins using cubic boxes of TIP3P water molecules in such a way that there was a 20 Å spacing between the protein and the edges. Following that, Na⁺ ions were added to neutralize the systems (8 for 4JQO and 7 for 3KZN), and the topology files and initial coordinates were produced using the force field ff14SB [40].

We carried out our molecular dynamic simulations using periodic boundary conditions, with H bond restrictions using the SHAKE algorithm [41] and 2 fs time step. The particle-mesh Ewald (PME) method was used to treat long-range electrostatic interactions with a 9 Å cutoff. The simulation runs were carried out in a multi-phase protocol, starting with 13,000 minimization using the conjugate gradient algorithm, followed by heating (NVT) and pressurization (NPT) phases using a Langevin thermostat [42], with equilibration values at 310 K and 1 atm, respectively. Then, the systems were equilibrated for 10 ns.

Simulations for the two proteins were carried out at three different temperatures: 310 K, 373 K, 436 K. We ran MD simulations in triplicate for the two proteins at each temperature, totaling 18 production trajectories of 50 ns each.

Analyses of RMSD, RMSF, and native contact fraction (Q), using the first frame as a reference, were used to study the equilibration and temporal stability of the systems. Principal component analysis (PCA) and a dynamic cross-correlation map (DCCM) analysis with the Bio3D tool [43] were performed to investigate the stability and flexibility of the proteins.

Results and discussion

The trajectories produced for knotted protein (PDB 3KZN) and unknotted (PDB 4JQO) allowed us to evaluate the structural stability against temperature changes. Besides the more conventional analysis performed, using RMSD, RMSF, and the fraction of native contacts, we also present data from the dynamic assessment of the structures by PCA and DCCM.

Molecular dynamic stability analysis

The MD trajectories were analyzed by using RMSD, RMSF, as well as the fraction of native contacts (Q). The data presented in Figs. 3 and 4 are average values for triplicates performed at each temperature studied. Since the RMSD values do not show peaks or great fluctuations along the trajectories, varying around 1 Å for temperatures of 310 K and 373 K, we have assumed that the production runs were carried out with equilibrated systems. At 436 K, as expected, the values obtained were higher, since the added kinetic energy makes the system jiggle more vigorously. Even at these high temperatures, however, there was no highlighted peak that showed complete destabilization in the trajectory.

At 310 K, it is possible to observe that for the knot moiety, residues 174 to 251 marked by the black rectangle, the fluctuation values are not very high in the knotted protein (PDB 3KZN in blue). As for the unknotted protein (PDB 4JQO in green), a high fluctuation value of 2.53 Å was observed for residue GLY243, which is in a loop moiety, well within the range that corresponded to the knot in the other system. Despite the lack of significant fluctuations inside the knot, it is still possible to observe high values of RMSF in the knotted protein, but they occur outside of the knot moiety. For instance, there is a region of large fluctuations between residues 86 and 88 (ASP86, GLY87, ASP88), but these are in a loop moiety, which justifies greater flexibility compared to the values highlighted in Fig. 3.

At the temperature of 373 K, even with the temperature rise, the knot moiety in 3KZN remained without large fluctuations, whereas the RMSF value for residue GLY243, in

the same region as the knot, was further increased (4.73 Å) for the 4JQO structure.

The highest RMSF values were observed for residues outside the knot moiety. Residue 243 (GLY243 highlighted) presents a fluctuation of 5.31 Å for the unknotted protein (orange line). The moieties outside the knot showed greater fluctuation for the knotted protein below the moiety containing the knot (between residues 1 and 173) and for the unknotted protein, above the moiety corresponding to the knot (residues above 251).

In Fig. 4, we plotted the fraction of native contacts (Q) of knotted and unknotted proteins, as a function of time, for the three temperatures. Comparing the Q profiles for the two proteins, we can see similar behavior at 310 K. The blue and green trajectories did not present a significant decrease of native contacts, with values above 0.95 along all the 50 ns (averaged values in triplicate).

At 373 K, both proteins show a small decrease in native contacts. Although the red curve corresponding to the knotted protein shows a slighter decrease, these values are very close to the black curve for the unknotted protein. Thus, we cannot yet state that the knotted protein displays greater stability at this temperature, as had been observed by Xu and collaborators [6], when they evaluated proteins considerably smaller than those used in our study. It is worth mentioning that, in that study, the protein was compared to itself after the handmade removal of one crossing of the chain that characterized the knot. In this study, we are evaluating two native structures with no geometric edition.

In order to assess the resistance to the decrease in native contacts, we raised the temperature to 436 K. The purple and orange curves represent the profiles for the 3KZN and the 4JQO proteins, respectively. We can then see in Fig. 4 that both structures suffered a significant decrease in their native contacts. However, over the 50 ns, the profiles did not show any distinct behavior between the two proteins.

Analysis on conformational behavior through PCA and DCCM

Principal component analysis was used to investigate the conformational behavior of the structures along the MD trajectories. Figures 5 and 6 show the PCA data for the structures at the three temperatures chosen.

Usually, the first two PCs are presented, which together can explain the major variance of the conformations visited during the simulation [44, 45]. For the knotted protein used in the present study, the total explained variance was approximately 45% for 310 K, 39% for 373 K, and 54% for 436 K. This information is summarized in Fig. 5.

For the two lowest temperatures, the red and blue conformations are practically aggregated, showing high stability of the structure over the 50 ns. Figure 5a presents a

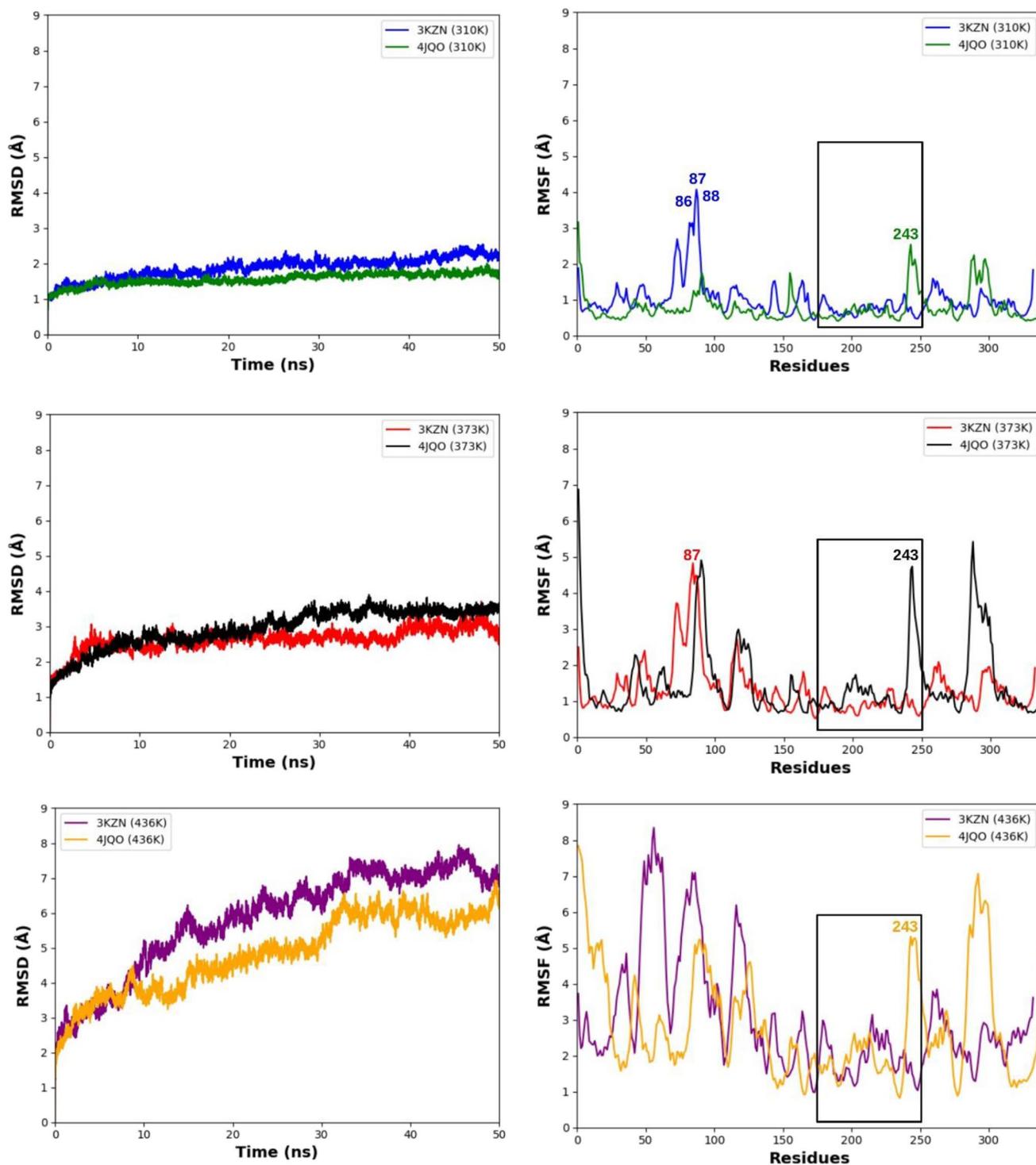


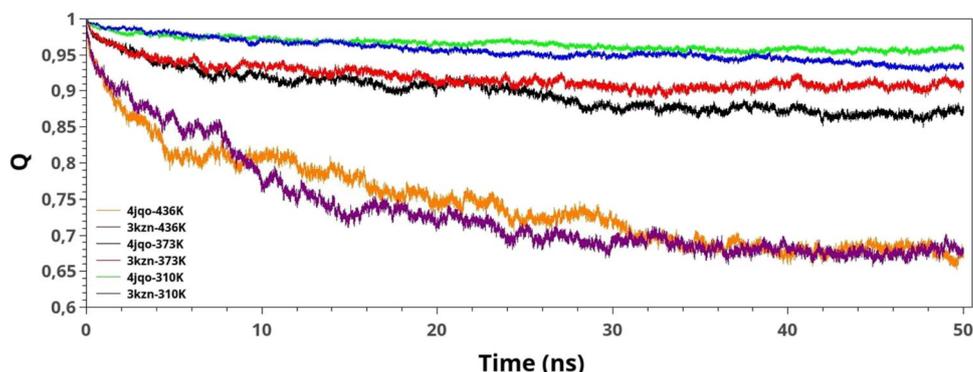
Fig. 3 RMSD and RMSF profiles of the knotted and unknotted proteins. The black square delimits the residues that compose the knot moiety. The values refer to the triplicate simulations

graph of residues versus PCs; there is no great highlight of any residue in relation to what appears in PC1 (black line) and PC2 (blue line). Thus, the motion described by the PCs has aggregated the results of all residues, with

no specific moiety contributing disproportionately to the variation.

At the temperature of 373 K, data presented in Fig. 5b, high stability is still observed, even with the increase in

Fig. 4 Native contacts (Q) over the 50 ns of MD simulation. A scheme of colors was used to represent the MD simulations performed at different temperatures: knotted protein (PDB 3KZN) in blue (310 K), red (373 K), and purple (436 K); unknotted protein (PDB 4JQO) in green (310 K), black (373 K), and orange (436 K). Values refer to triplicate



temperature for the protein under study. The PCA graph shows the very concentrated data for the red and blue conformations. The graph of residues versus PCs, at this temperature, shows a steep rise between residues 70 and 90 for PC1 and PC2, indicating that this moiety has made the greatest contribution to the variance observed in the PCA.

Although a greater contribution of residues from the moiety between 70 and 90 to PCA is observed, these did not affect the distribution of conformations. There is still a high density, as indicated by the elevated concentration of red and blue dots, which points to high stability even at a temperature of 373 K.

For the highest temperature of 436 K, there is a change in relation to the others, as seen in Fig. 5c. Throughout the distribution in the PCA, a greater dispersion between red and blue is observed, which shows that, for this temperature, the protein has lost some of its stability. This is consistent with what has been observed in other studies, where the stability of knotted protein was assessed at a very high temperature [6].

According to the graph of residues versus PCs, there is a contribution of other residues to PC1 and PC2 when compared to the other two temperatures of the study. However, the moiety of greatest influence for the variance observed in the PCA is between residues 50 and 70 for PC1 and 60 and 80 for PC2, which is near the highest contributions observed at other temperatures.

The PCA data for the unknotted protein is shown in Fig. 6. For the three temperatures, the behavior is quite different when compared to the structure containing the knot shown in Fig. 5. Considering the first two PCs responsible for the largest data variance, the percentage of explained variance was around 37% for 310 K and 373 K, and 43% at 436 K.

At 310 K, Fig. 6a, in contrast to what is observed for the knotted protein, there is a concentration of intermediate conformations (white dots), which points to a less stability when compared to the same temperature in Fig. 5a, where red and blue conformations appear dense with few white conformations. At this temperature, the amplitude of

the contribution of the residues to the PCs (residue position \times PCs) is considerably greater for the unknotted than that for the knotted protein.

In addition, several residues stand out with a greater contribution to PC1 (black line) and PC2 (blue line), as we can see in the graph of residues versus PCs. An interesting point is the highlight for the residue close to 250 that is in the moiety corresponding to the knot moiety in 3KZN. This is not observed in Fig. 5a, at the temperature of 310 K. This corroborates with what was observed for the RMSF, where residue GLY243 was shown to have the highest flexibility.

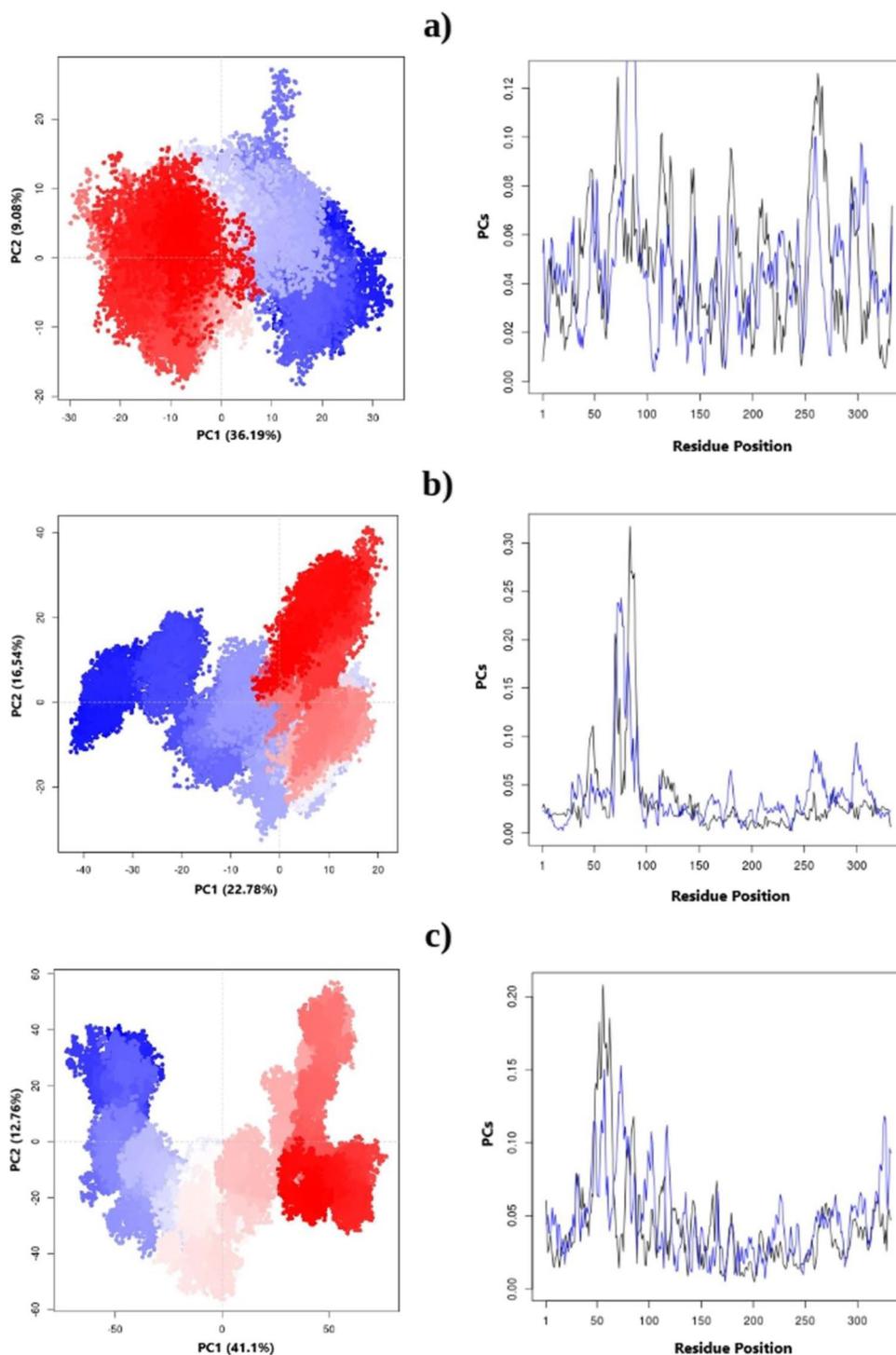
At 373 K, the PCA showed a high density of red and blue conformations, similar to what we observed at this temperature for the knotted protein. However, when evaluating the contributions to the behavior of the PCA through the graph of residues versus PCs, differences emerge. For PC1 (black line), the greatest contribution is from residues around 300, which was not observed for the knotted protein. PC2 (in blue) has two significant peaks, the higher at around residue 90 and a lower one near residue 250.

At 436 K, Fig. 6c, the distribution of conformations with greater amounts in red is clear, along PC1. There are more red dots at 436 K, compared to the knotted protein, Fig. 5c. There are groups of conformations (in red) between 40 and 60 in the PCA graph, and a smaller one between 20 and 40, and yet another approximately between -10 and 10. This shows groups of variable conformations for the structure at this temperature, where there was a greater variety of conformations.

The most prominent residues in the PCs are the same observed in 373 K; however, PC2 in blue has peaks at around 300 and just shy of the 250 residue positions. These are the moieties of residues that contributed the most to the observed variance at the PCA.

The correlated behavior of the movement of residues for the two proteins is shown in Fig. 7. The correlation is presented as a series of DCCM graphs. We present, side by side, the data for the three temperatures considered: a and b at 310 K, c and d at 373 K, e and f at 436 K. DCCM has

Fig. 5 PCA of the MD simulations (knotted protein, PDB 3KZN). **a** 310 K; **b** 373 K; **c** 436 K. The contributions of the residuals for PC1 in black and PC2 in blue are shown alongside



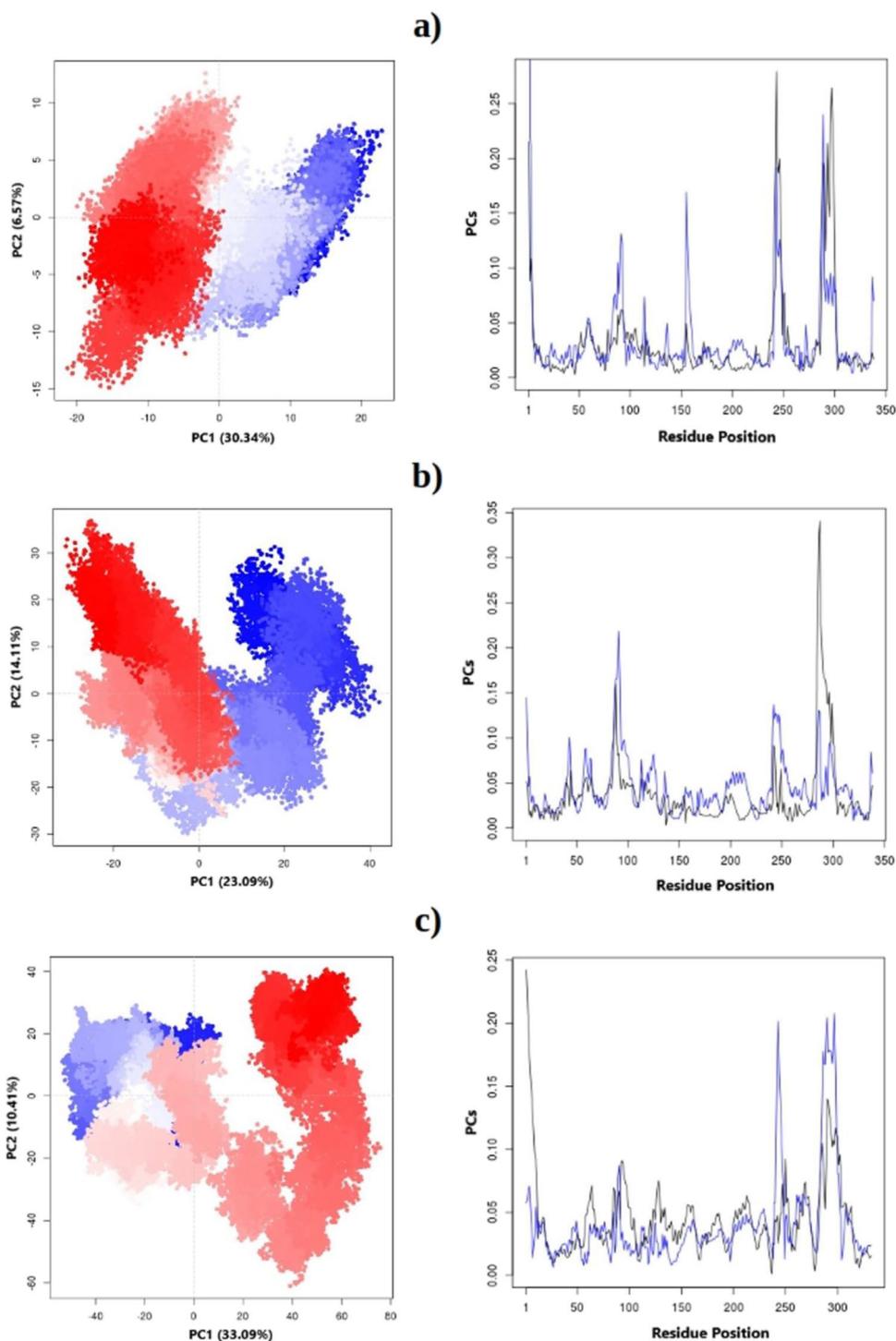
been used successfully previously in a comparative analysis of other systems [44, 45].

All data on the right in Fig. 7 are for the knotted protein and on the left for the unknotted. The behaviors of both proteins across the temperature range are so distinct that even a cursory visual inspection in the graph is enough to notice it. The stronger the cyan color (positive values), the stronger the correlation in the motion of the residues expressed in the

DCCM graph, showing that the residues move in the same direction. Otherwise, where the pink color prevails (negative values), the movements have gone in the opposite direction, showing anti-correlation. The rectangles delimit the moieties of significant change observed with the rise in temperature.

In Fig. 7a and b (unknotted and knotted at the lowest study temperature, respectively), the difference in behavior between the structures is quite clear. At 310 K, we see that

Fig. 6 PCA of the MD simulations (unknotted protein, PDB 4JQO). **a** 310 K; **b** 373 K; **c** 436 K. The contributions of the residuals for PC1 in black and PC2 in blue are shown alongside



there is a low correlation of the residue movement for the unknotted protein. There are only small areas in light pink in the matrix shown in Fig. 7a, showing anti-correlation movements (moieties between 240 and 250 and between 280 and 300 with the moiety from 50 to 120). Contrary to what was observed for the unknotted protein with a small moiety of anti-correlation movements, for the knotted protein in the DCCM graph in Fig. 7b, there is a pink color with intense

tones (values close to -1), pointing to high anti-correlation (that is, moieties of opposite movements) for the moieties of 175 to 220 and 250 to 300, with the moiety between residues 45 and 150. There is also, for the knotted protein, high correlation (in cyan) between the moiety of residues 250 and 275, with the moiety 175 to 225 that is inserted in the knot.

With the temperature rise to 373 K, the unknotted protein showed an increase in its movements, as we can see in

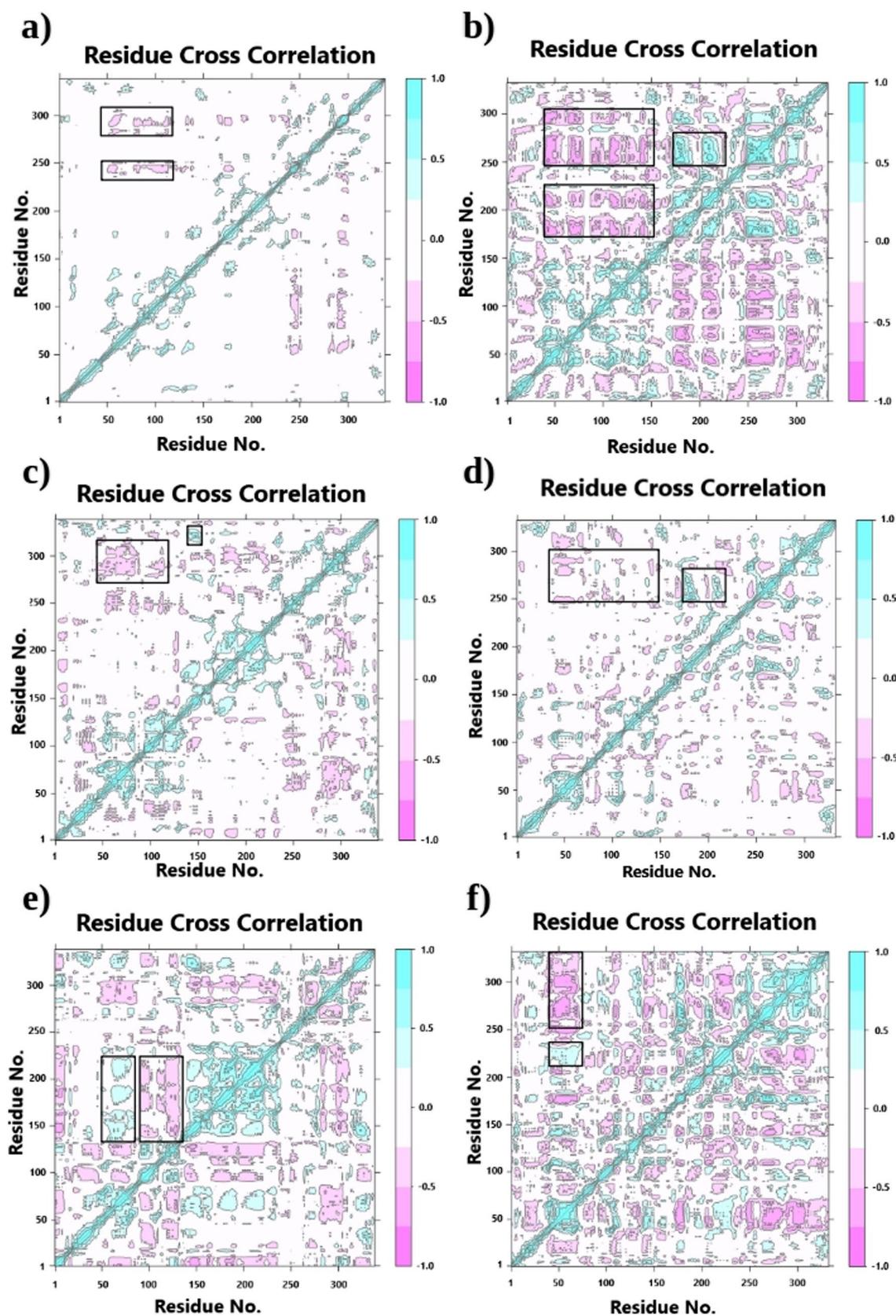


Fig. 7 The DCCM of the unknotted protein PDB 4JQO (**a**, **c**, **e**) and knotted protein PDB 3KZN (**b**, **d**, **f**) at temperatures 310 K (top), 373 K (middle), and 436 K (bottom). Regions of significant differences for correlation and anti-correlation shown in the black squares

Fig. 7c compared to a. There is a considerable increase in anti-correlation movements in the graph shown in Fig. 7c, as we can see by the increase in the pink area when comparing the moiety of residues 275 and 315 with the moiety between 50 and 120. The correlation of the movement of the residues (values in cyan) appears in some regions to be the same as the increase observed in the moiety above 315 with the moiety between 145 and 155. That is to say, there was an increase in the region formed between the residues above 315 and those in the region between 145 and 155, when comparing information from Fig. 7a and c.

For the knotted protein compared with the unknotted, in 373 K, a different behavior can be seen, since there is a decrease in movements. In Fig. 7d, there is a very sharp decrease in the correlation and anti-correlation movements. And there is a decrease in both types of movements when the temperature for the knotted protein is raised. For example, the anti-correlation of the moiety of residue 250 to 300 with the moiety of 45 to 150 is noticeably reduced. The knot moiety has a correlation movement observed between the residues 250 and 275 with the region 175 and 225 (cyan area in the black box) that decreased with the higher temperature; but visually less reduction occurs when compared to the reduction occurred for the cited anti-correlation moieties.

For the highest temperature (436 K), we noticed a significant increase in both correlation and anti-correlation movements for both proteins, when compared to the 373 K temperature represented in Fig. 7c and d, for the unknotted and knotted proteins, respectively. This suggests that both movements of the structures were affected by the temperature rise, resulting in a noticeable change in the DCCM graphs as shown in Fig. 7e and f.

The 436 K behavior for the unknotted protein, as seen in Fig. 7e, stands out mainly for the appearance of correlation movements in the moiety comprising residues between 135 and 225, with the moiety between 60 and 80, which did not happen at lower temperatures. And there was also an increase in anti-correlation, with emphasis on the one between segments 135–225 and 90–130.

For the knotted protein, presented in Fig. 7f, there is a significant increase in anti-correlation movements that have already been appeared at 373 K, but which appear even stronger at 436 K, for example, between segment 250–332 and 40–70. Correlation motion also increased overall, for instance, the correlation between segments 215–230 and 40–70 showed a considerable increase when compared to lower temperatures.

PCA and DCCM data showed greater resistance to conformational changes for 3KZN compared to 4JQO. Our study showed that a knotted protein is more stable under temperature conditions considered. This result agrees with that obtained by Xu and collaborators [46], when they

analyzed the importance of the knot in a protein. They observed reduced stability for the unknotted one. In the experimental study carried out by Sirarimoju and collaborators [35], knotted proteins were also considered as more resistant to the changes imposed on the systems.

Conclusions

In this work, the structural dynamics of a knotted protein was investigated in comparison to its unknotted analog using molecular dynamics at three temperatures.

RMSF showed that at all evaluated temperatures, the two proteins had their more flexible moieties outside the knot. The greatest fluctuations were observed for the unknotted protein (PDB 4JQO), which showed a prominent fluctuation for residue 243, within the region equivalent to the knot moiety in the knotted protein. The fluctuation was further enhanced when temperature was increased to 373 K and 436 K for both structures. The fraction of native contacts showed a significant decrease only at 436 K for the two proteins.

PCA showed that the unknotted protein had access to a more varied set of conformations, which may show less resistance to conformational changes compared with the knotted protein. The DCCM data showed that the unknotted protein underwent major changes in its movements (correlation and anti-correlation) related to the temperature changes, which suggests that it is prone to greater conformational variations, indicating less stability.

The data obtained corroborate other studies in the literature that linked greater stability to a knotted protein compared to the unknotted analog. Highlighted in this study, unlike what has been done before, was the use of structures without geometric modifications for comparison of the knotted protein structural dynamics.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00894-022-05094-y>.

Acknowledgements The authors acknowledge the physical structure and computational support provided by Universidade Federal da Paraíba (UFPB), the computer resources of Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP). The English text of this paper has been revised by Sidney Pratt, Canadian, MAT (The Johns Hopkins University), RSAdip—TESL (Cambridge University).

Author contribution José Cícero Alves Silva: wrote the original manuscript, performed research, curation, and data analysis. Elton José Ferreira Chaves: wrote the original manuscript, curation and data analysis. Gabriel Aires Urquiza de Carvalho: coordinator and revised the original manuscript. Gerd Bruno Rocha: coordinator, planned the research, fundraising, revised the original manuscript. All the authors have read and approved the final manuscript.

Funding This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) through the research project Bioinformática Estrutural de Proteínas: Modelos, Algoritmos e Aplicações Biotecnológicas (Edital Biologia Computacional 51/2013, processo AUXPE1375/2014 da CAPES). G.B.R. received support from the Brazilian National Council for Scientific and Technological Development (CNPq grant no. 309761/2017–4).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Declarations

Competing interests The authors declare no competing interests.

References

- Richardson JS (1977) β -Sheet topology and the relatedness of proteins. *Nature* 268:495–500. <https://doi.org/10.1038/268495a0>
- Mansfield ML (1994) Are there knots in proteins? *Nat Struct Biol* 1:213–214. <https://doi.org/10.1038/nsb0494-213>
- Sułkowska JI, Rawdon EJ, Millett KC et al (2012) Conservation of complex knotting and slipknotting patterns in proteins. *Proc Natl Acad Sci* 109:E1715–E1723. <https://doi.org/10.1073/pnas.1205918109>
- Dabrowski-Tumanski P, Stasiak A, Sulowska JI (2016) In search of functional advantages of knots in proteins. *PLoS ONE* 11:1–14. <https://doi.org/10.1371/journal.pone.0165986>
- Dabrowski-Tumanski P, Sulowska JI (2017) Topological knots and links in proteins. *Proc Natl Acad Sci* 114:3415–3420. <https://doi.org/10.1073/pnas.1615862114>
- Xu Y, Li S, Yan Z et al (2018) Stabilizing effect of inherent knots on proteins revealed by molecular dynamics simulations. *Biophys J* 115:1681–1689. <https://doi.org/10.1016/j.bpj.2018.09.015>
- Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
- Perlinska AP, Stasiulewicz A, Nawrocka EK et al (2020) Restriction of S-adenosylmethionine conformational freedom by knotted protein binding sites. *PLoS Comput Biol* 16:e1007904. <https://doi.org/10.1371/journal.pcbi.1007904>
- Zhao Y, Dabrowski-Tumanski P, Niewieczeral S, Sulowska JI (2018) The exclusive effects of chaperonin on the behavior of proteins with 52 knot. *PLoS Comput Biol* 14:e1005970. <https://doi.org/10.1371/journal.pcbi.1005970>
- Sułkowska JI, Noel JK, Onuchic JN (2012) Energy landscape of knotted protein folding. *Proc Natl Acad Sci* 109:17783–17788. <https://doi.org/10.1073/pnas.1201804109>
- Wang L-W, Liu Y-N, Lyu P-C et al (2015) Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of *Helicobacter pylori*. *J Phys Condens Matter* 27:354106. <https://doi.org/10.1088/0953-8984/27/35/354106>
- Flapan E, He A, Wong H (2019) Topological descriptions of protein folding. *Proc Natl Acad Sci USA* 116:9360–9369. <https://doi.org/10.1073/pnas.1808312116>
- Sułkowska JI (2020) On folding of entangled proteins: knots, lassos, links and θ -curves. *Curr Opin Struct Biol* 60:131–141. <https://doi.org/10.1016/j.sbi.2020.01.007>
- Begun A, Liubimov S, Molochkov A, Niemi AJ (2021) On topology and knotty entanglement in protein folding. *PLoS ONE* 16:e0244547. <https://doi.org/10.1371/journal.pone.0244547>
- Barbensi A, Goundaroulis D (2021) f-distance of knotoids and protein structure. *Proc R Soc A* 477:20200898. <https://doi.org/10.1098/rspa.2020.0898>
- Sriramoju MK, Chen Y, Lee Y-TC, Hsu S-TD (2018) Topologically knotted deubiquitinases exhibit unprecedented mechanostability to withstand the proteolysis by an AAA+ protease. *Sci Rep* 8:1–9. <https://doi.org/10.1038/s41598-018-25470-0>
- San Martín Á, Rodríguez-Aliaga P, Molina JA et al (2017) Knots can impair protein degradation by ATP-dependent proteases. *Proc Natl Acad Sci* 114:9864–9869. <https://doi.org/10.1073/pnas.1705916114>
- Wojciechowski Michał and Gómez-Sicilia À, Carrión-Vázquez M, Cieplak M (2016) Unfolding knots by proteasome-like systems: simulations of the behaviour of folded and neurotoxic proteins. *Mol Biosyst* 12:2700–2712. <https://doi.org/10.1039/C6MB00214E>
- Fonseka HYY, Javidi A, Oliveira LFL et al (2021) Unfolding and translocation of knotted proteins by Clp biological nanomachines: synergistic contribution of primary sequence and topology revealed by molecular dynamics simulations. *J Phys Chem B* 125:7335–7350. <https://doi.org/10.1021/acs.jpbc.1c00898>
- Sivertsson EM, Jackson SE, Itzhaki LS (2019) The AAA+ protease ClpXP can easily degrade a 31 and a 52-knotted protein. *Sci Rep* 9:1–14. <https://doi.org/10.1038/s41598-018-38173-3>
- Sriramoju MK, Chen Y, Hsu S-TD (2020) Protein knots provide mechano-resilience to an AAA+ protease-mediated proteolysis with profound ATP energy expenses. *Biochim Biophys Acta (BBA)-Proteins Proteomics* 1868:140330. <https://doi.org/10.1016/j.bbapap.2019.140330>
- Lim NCH, Jackson SE (2015) Molecular knots in biology and chemistry. *J Phys Condens Matter* 27:354101. <https://doi.org/10.1088/0953-8984/27/35/354101>
- Virnau P, Mallam A, Jackson S (2010) Structures and folding pathways of topologically knotted proteins. *J Phys Condens Matter* 23:33101. <https://doi.org/10.1088/0953-8984/23/3/033101>
- Faisca PFN (2015) Knotted proteins: a tangled tale of structural biology. *Comput Struct Biotechnol J* 13:459–468. <https://doi.org/10.1016/j.csbj.2015.08.003>
- Potestio R, Micheletti C, Orland H (2010) Knotted vs. unknotted proteins: evidence of knot-promoting loops. *PLoS Comput Biol* 6:e1000864. <https://doi.org/10.1371/journal.pcbi.1000864>
- Forgan RS, Sauvage J-P, Stoddart JF (2011) Chemical topology: complex molecular knots, links, and entanglements. *Chem Rev* 111:5434–5464. <https://doi.org/10.1021/cr200034u>
- Jackson SE, Suma A, Micheletti C (2017) How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Curr Opin Struct Biol* 42:6–14. <https://doi.org/10.1016/j.sbi.2016.10.002>
- Dabrowski-Tumanski P, Rubach P, Goundaroulis D et al (2019) KnotProt 2.0: a database of proteins with knots and other entangled structures. *Nucleic Acids Res* 47:D367–D375. <https://doi.org/10.1093/nar/gky1140>
- Beccara SA, Škrbić T, Covino R et al (2013) Folding pathways of a knotted protein with a realistic atomistic force field. *PLoS Comput Biol* 9:e1003002. <https://doi.org/10.1371/journal.pcbi.1003002>
- Sułkowska JI, Sulkowski P, Onuchic J (2009) Dodging the crisis of folding proteins with knots. *Proc Natl Acad Sci* 106:3119–3124. <https://doi.org/10.1073/pnas.0811147106>
- Noel JK, Onuchic JN, Sulowska JI (2013) Knotting a protein in explicit solvent. *J Phys Chem Lett* 4:3570–3573. <https://doi.org/10.1021/jz401842f>

32. Perlinska AP, Kalek M, Christian T et al (2020) Mg²⁺-dependent methyl transfer by a knotted protein: a molecular dynamics simulation and quantum mechanics study. *ACS Catal* 10:8058–8068. <https://doi.org/10.1021/acscatal.0c00059>
33. Bölinger D, Sułkowska JI, Hsu H-P et al (2010) A Stevedore's protein knot. *PLoS Comput Biol* 6:e1000731. <https://doi.org/10.1371/journal.pcbi.1000731>
34. Wu X, Xu P, Wang J, et al (2015) Folding mechanisms of trefoil knot proteins studied by molecular dynamics simulations and Gomodel. In: *Advance in Structural Bioinformatics*. Springer, pp 93–110. https://doi.org/10.1007/978-94-017-9245-5_8
35. Sriramoju MK, Yang T-J, Hsu S-TD (2018) Comparative folding analyses of unknotted versus trefoil-knotted ornithine transcarbamylases suggest stabilizing effects of protein knots. *Biochem Biophys Res Commun* 503:822–829. <https://doi.org/10.1016/j.bbrc.2018.06.082>
36. Dabrowski-Tumanski P, Stasiak A, Sulkowska JI (2016) In search of functional advantages of knots in proteins. *PLoS ONE* 11:e0165986. <https://doi.org/10.1371/journal.pone.0165986>
37. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802. <https://doi.org/10.1002/jcc.20289>
38. Case DA, Ben-Shalom IY, Brozell SR et al (2018) AMBER 2018; 2018. University of California, San Francisco
39. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB-2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665–W667. <https://doi.org/10.1093/nar/gkh381>
40. Maier JA, Martinez C, Kasavajhala K et al (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
41. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
42. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys* 103:4613–4621. <https://doi.org/10.1063/1.470648>
43. Grant BJ, Rodrigues APC, ElSawy KM et al (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22:2695–2696. <https://doi.org/10.1093/bioinformatics/btl461>
44. Zhou L, Ma Y-C, Tang X et al (2021) Identification of the potential dual inhibitor of protein tyrosine phosphatase sigma and leukocyte common antigen-related phosphatase by virtual screen, molecular dynamic simulations and post-analysis. *J Biomol Struct Dyn* 39:45–62. <https://doi.org/10.1080/07391102.2019.1705913>
45. Wei-Ya L, Yu-Qing D, Yang-Chun M et al (2019) Exploring the cause of the inhibitor 4AX attaching to binding site disrupting protein tyrosine phosphatase 4A1 trimerization by molecular dynamic simulation. *J Biomol Struct Dyn* 37:4840–4851. <https://doi.org/10.1080/07391102.2019.1567392>
46. Xu Y, Li S, Yan Z et al (2019) Revealing cooperation between knotted conformation and dimerization in protein stabilization by molecular dynamics simulations. *J Phys Chem Lett* 10:5815–5822. <https://doi.org/10.1021/acs.jpcclett.9b02209>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.