Journal of Health Informatics

J Health Inform. 2026;18:e1822
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1822

ARTIGO COMPLETO
CBIS 2026

DALI e TM-align: abordagens complementares para alinhamento estrutural de proteínas e sua relevância em oncologia

DALI and TM-align: complementary approaches to protein structural alignment and their relevance in oncology

Samuel Henrique de Moura Rodrigues[1] ✉, Ana Paula Roque da Silva[1], Filipe Pessoa Sousa[1], Helainy Ignácio de Almeida Torres[1], Leonardo de Almeida Cavadas[1], Alexandre da Costa Sena[1]

ORCID: 0009-0003-0744-3154 · 0000-0002-9076-0091 · 0009-0003-6356-0013 · 0000-0001-8529-6645 · 0000-0002-4747-2246 · 0000-0002-8931-9397

[1] Universidade do Estado do Rio de Janeiro, Rio de Janeiro (RJ), Brasil.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1822

DOI
10.59681/2175-4411.v18.iEspecial.2026.1822

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Samuel Henrique de Moura Rodrigues
shrodrigues_mat@outlook.com

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Objetivo: Realizar uma comparação abrangente entre os algoritmos DALI e TM-Align, avaliando a complementaridade entre esses algoritmos para análise estrutural de proteínas ligadas ao câncer. **Método:** Alinhamento estrutural de proteínas, obtidas no Protein Data Bank (PDB), através das ferramentas DALI e TM-Align, e comparação e análise dos resultados obtidos. **Resultados:** Os algoritmos DALI e TM-Align possuem características complementares fundamentais para a análise estrutural de proteínas associadas ao câncer. Embora ambos tenham indicado baixa similaridade estrutural global entre as proteínas analisadas, cada método contribuiu de forma distinta para a interpretação dos alinhamentos. **Conclusão:** Abordagens integradas de alinhamento estrutural contribuem para uma compreensão mais abrangente dos mecanismos moleculares envolvidos na oncogênese, auxiliando na interpretação funcional de alterações estruturais e no potencial desenvolvimento de novos alvos terapêuticos.

Descritores: Alinhamento de proteínas; DALI; TM-Align.

ABSTRACT

Objective: To perform a comprehensive comparison between the DALI and TM-Align algorithms, evaluating the complementarity between these algorithms for the structural analysis of cancer-related proteins. **Method:** Structural alignment of proteins obtained from the Protein Data Bank (PDB) using the DALI and TM-Align tools, followed by comparison and analysis of the results obtained. **Results:** The DALI and TM-Align algorithms possess fundamental complementary characteristics for the structural analysis of cancer-associated proteins. Although both indicated low overall structural similarity between the analyzed proteins, each method contributed distinctly to the interpretation of the alignments. **Conclusion:** Integrated structural alignment approaches contribute to a more comprehensive understanding of the molecular mechanisms involved in oncogenesis, aiding in the functional interpretation of structural alterations and the potential development of new therapeutic targets.

Keywords: Protein alignment; DALI; TM-Align.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: Rodrigues SHM, da Silva APR, Sousa FP, Torres HIA, Cavadas LA, Sena AC. DALI e TM-align: abordagens complementares para alinhamento estrutural de proteínas e sua relevância em oncologia. J Health Inform 2026;18:e1822. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1822

1 INTRODUÇÃO

O alinhamento estrutural de proteínas é uma ferramenta essencial na bioinformática, desempenhando um papel central na compreensão de relações evolutivas, funções biológicas e mecanismos moleculares. Estruturas tridimensionais de proteínas frequentemente compartilham similaridades, mesmo quando suas sequências apresentam baixa identidade, o que torna o alinhamento estrutural crítico para revelar tais correspondências e identificar regiões conservadas ou funcionalmente relevantes [1]. Esses métodos são amplamente utilizados para desvendar alterações estruturais que estão frequentemente associadas a doenças, como mutações em tumores que afetam processos biológicos essenciais.

Entre os diversos algoritmos para alinhamento estrutural, dois se destacam amplamente: o DALI e o TM-Align. Esses métodos diferem em suas abordagens e são amplamente utilizados em estudos comparativos. O DALI, introduzido por Holm [2] e Sander [3], baseia-se na comparação de matrizes de distância, utilizando relações espaciais entre átomos para identificar similaridades estruturais. Sua abordagem é particularmente eficaz na detecção de relações evolutivas distantes entre proteínas. Em contraste, o TM-Align, desenvolvido por Zhang e Skolnick [4], utiliza o TM-score, uma métrica que maximiza a sobreposição estrutural global, priorizando regiões inteiras bem alinhadas e, portanto, é frequentemente considerado mais robusto em termos de avaliação quantitativa de similaridade estrutural.

A escolha do algoritmo adequado é influenciada por vários fatores, como a sensibilidade às similaridades distantes, a precisão na identificação de regiões críticas e a relevância funcional dos alinhamentos gerados. No contexto do câncer, onde alterações em regiões estruturais podem impactar profundamente a funcionalidade proteica, essas considerações são cruciais. Sítios de ligação ou interfaces de interação proteína-proteína são exemplos de regiões cuja conservação ou modificação pode determinar mecanismos de resistência a drogas ou propagação tumoral. Assim, a capacidade de identificar tais regiões é um critério-chave na avaliação de algoritmos de alinhamento.

Um aspecto igualmente importante é a avaliação quantitativa da qualidade do alinhamento. Métricas robustas e objetivas permitem validar a adequação dos resultados e garantir que as similaridades detectadas sejam biológica e funcionalmente relevantes. Tais técnicas não apenas ajudam a comparar algoritmos, mas também fornecem um contexto funcional para entender como alterações estruturais impactam processos biológicos.

DALI e TM-Align possuem suas respectivas vantagens e limitações. O DALI é amplamente reconhecido por sua capacidade de detectar relações distantes, tornando-o uma escolha comum em estudos evolutivos. Por outro lado, o TM-Align é frequentemente preferido em aplicações que exigem alinhamentos precisos e quantitativos, como na modelagem de estruturas proteicas ou na avaliação de modelos preditivos. Estudos de *benchmarking*, como os conduzidos por Zhang e Skolnick [4], demonstraram que o TM-Align supera o DALI em termos de velocidade e precisão ao comparar estruturas com baixa identidade de sequência. Além disso, a métrica TM-score fornece uma interpretação intuitiva da qualidade do alinhamento, sendo diretamente proporcional à quantidade de regiões bem alinhadas.

No contexto de tumores, onde mutações frequentemente alteram regiões estruturais críticas, como domínios catalíticos ou áreas de interação, a avaliação funcional dos alinhamentos torna-se ainda mais relevante. Regiões que são altamente conservadas em proteínas homólogas podem fornecer pistas cruciais sobre mecanismos de ação ou resistência. Por exemplo, alterações em sítios de ligação podem modificar a afinidade por ligantes, influenciando diretamente a eficácia de terapias baseadas em pequenas moléculas. Assim, algoritmos que conseguem identificar essas regiões com alta confiabilidade são ferramentas indispensáveis na pesquisa biomédica.

Nesse contexto, este estudo tem como objetivo realizar uma comparação abrangente entre os algoritmos DALI e TM-Align, com foco em três aspectos principais: sensibilidade na detecção de similaridades estruturais distantes, precisão na identificação de regiões funcionalmente relevantes e relevância biológica dos alinhamentos gerados. Para tanto, serão utilizados estudos de caso envolvendo estruturas proteicas relacionadas ao câncer e *benchmarks* quantitativos de desempenho. A integração dessas abordagens permitirá uma análise detalhada das capacidades e limitações de cada algoritmo, fornecendo subsídios para sua aplicação em contextos específicos.

Os parágrafos a seguir descrevem trabalhos relacionados relevantes encontrados na literatura. O artigo apresentado em [5] mostra os avanços e armadilhas no alinhamento estrutural de proteínas. De acordo com o artigo, o objetivo do alinhamento estrutural pode ser definido geometricamente, como a sobreposição de um número máximo de pontos com desvios espaciais mínimos, ou geneticamente, como a identificação de resíduos homólogos, destacando que métodos de comparação estrutural que permitem flexibilidade e plasticidade geram os alinhamentos biologicamente mais significativos.

Por outro lado, a importância de encontrar alinhamentos estruturais biologicamente significativos levou ao intenso desenvolvimento de métodos para gerar alinhamentos e avaliar sua qualidade. No entanto, esses métodos produzem resultados

conflitantes e nenhum foi aceito como claramente superior. Assim, o artigo em [6] descreve uma medida de qualidade de alinhamento, o valor-I, que utiliza o conteúdo de informação de mensagens que comprimem sem perda os átomos Cα de um par de estruturas proteicas, dado um alinhamento proposto. Um valor-I menor indica um alinhamento superior. O método não possui parâmetros ajustáveis, pois é baseado no princípio bayesiano formal de inferência de comprimento mínimo de mensagem.

Não foram encontrados trabalhos comparando os alinhadores DALI e TM-Align, especialmente destacando a complementaridade de se utilizar os dois alinhadores para produzir resultados mais significativos. A Seção 2 descreve a metodologia adotada neste trabalho. Por sua vez, na Seção 3, são apresentados os resultados do trabalho, nos quais ambos os algoritmos foram empregados de forma complementar para analisar similaridades estruturais entre proteínas associadas a processos oncogênicos. Na Seção 4, realiza-se a discussão dos resultados, evidenciando os contextos em que o uso de cada algoritmo se mostra mais adequado. Por fim, na Seção 5, são apresentadas as conclusões.

2 MÉTODOS

O alinhamento de proteínas é uma tarefa de extrema importância na bioinformática, pois permite a identificação de estruturas semelhantes entre as proteínas. Essa informação é fundamental para entender a função, evolução e classificação de proteínas. A metodologia adotada neste trabalho consiste no alinhamento estrutural de proteínas relacionadas ao câncer, utilizando os alinhadores TM-Align e o DALI, seguido da comparação e análise dos resultados. As proteínas utilizadas neste trabalho foram obtidas do banco de dados *Protein Data Bank* (PDB), um repositório amplamente utilizado na bioinformática. Nas próximas subseções são apresentados os dois algoritmos de alinhamento empregados nesse trabalho (subseções 2.1 e 2.2), assim como as proteínas utilizadas neste estudo (subseção 2.3).

2.1 TM-Align

O algoritmo TM-Align (*Template Modeling Alignment*) tem como principal característica a precisão e velocidade no alinhamento das proteínas. Ele combina a matriz de rotação TM-*score* com programação dinâmica para identificar o melhor alinhamento estrutural entre um par de proteínas. Ele apresenta velocidade de alinhamento expressivamente maior que o DALI, o que o torna ideal para análises de grandes conjuntos de dados [4]. Em relação à precisão, esse algoritmo produz alinhamentos com maior acurácia e cobertura quando comparado com outros métodos e, além disso, apresenta disponibilidade de uso tanto on-line (através do site https://aideepmed.com/TM-align/), como também através da instalação do *software* de alinhamento.

Esse algoritmo apresenta uma forma de medir a semelhança entre as estruturas das proteínas chamada TM-score. Ele quantifica a similaridade entre duas estruturas, considerando tanto a distância entre os resíduos alinhados quanto o tamanho das proteínas. Após a realização do alinhamento das proteínas, o TM-Align retorna informações sobre as similaridades das estruturas através das métricas TM-score e o RMSD (*Root Mean Square Deviation*).

O TM-score é uma métrica calculada após se fazer o alinhamento de duas estruturas com o objetivo de minimizar a distância entre os resíduos correspondentes [4], onde para cada par de resíduos alinhados, calcula-se a distância entre seus átomos C_α, ou seja, os átomos de carbono alfa, que é um termo utilizado no contexto de bioquímica e de análise estrutural para indicar que um átomo de carbono ocupa a posição central de um aminoácido na estrutura primária de uma proteína. Em seguida, é feita a normalização dessas distâncias com base no comprimento de cada proteína analisada, para que o resultado (*score*) seja um valor entre 0 e 1. Logo, em cada comparação são fornecidos dois *scores* que diferem entre si pelo tamanho da proteína alvo. Quanto mais próximo o *score* for de 1, mais alta a similaridade estrutural entre as proteínas. Valores próximos a 0 indicam que as estruturas têm pouca ou nenhuma relação. A fórmula do TM-score é apresentada como:

$$TM\text{-}score = max\left[\frac{1}{L_{\text{alvo}}}\sum_{i=1}^{L_{\text{ali}}}\left(1+\frac{d_i}{d_0(L_{\text{alvo}})}\right)^{-2}\right] \quad \text{(a)}$$

onde L_{alvo} é o comprimento da proteína principal (alvo), L_{ali} é o número de resíduos alinhados e d_i é a distância entre o i-ésimo par de resíduos alinhados. O último desses parâmetros é o d_0, um fator de normalização baseado na distância média esperada entre resíduos em um alinhamento estrutural aleatório. Esse fator é definido pela expressão

$$d_0(L_{\text{target}}) = 1.24\sqrt[3]{L_{\text{target}}} - 1.8. \quad \text{(b)}$$

Assim como no DALI, a métrica RMSD também mede a diferença média entre os átomos correspondentes após o alinhamento estrutural. Outra informação fornecida pelo TM-Align é a L_{ali} (*Aligned length*), a quantidade de resíduos alinhados. Quanto mais alta essa métrica, melhor é o alinhamento das proteínas.

Esse algoritmo tem várias aplicações, dentre as quais podemos citar:

- Classificação de estruturas através da identificação de novas dobras proteicas;

- Modelagem de proteínas, ou seja, comparação de modelos preditos com estruturas encontradas de forma experimental;
- Análise funcional, por meio da qual são identificadas proteínas com funções semelhantes com base em sua estrutura.

O exemplo da Figura 1 traz o alinhamento par a par das proteínas teste 101M-A e 1MBA. Após o alinhamento, foi possível obter que, dos 154 resíduos da proteína 101M-A e dos 146 resíduos da proteína 1MBA, 143 resíduos foram alinhados. Isso significa que essas proteínas são semelhantes. O RMSD foi de 1,83, o que sugere que as duas proteínas têm estruturas altamente similares.

Foram calculados dois valores para o TM-score. O TM-score normalizado pelo comprimento de 101M-A foi de 0,81453 e TM-score normalizado pelo comprimento de 1MBA foi de 0,85377. Ambos os valores, próximos de 1, indicam que as duas proteínas apresentam a mesma dobra estrutural. Todas essas métricas corroboram a sobreposição das proteínas mostradas na Figura 1.

Figura 1. Sobreposição das proteínas 101M-A e 1MBA



Fonte: Elaborado pelos autores (2026).

2.2 Dali

O DALI (*Distance-matrix Alignment*) é um algoritmo conhecido por sua versatilidade no alinhamento das estruturas proteicas. Ele utiliza uma abordagem baseada em distâncias para comparar essas estruturas, o que permite a identificação de semelhanças locais e globais, mesmo em proteínas com diferentes topologias, isto é, diferentes funções e formas. Esse algoritmo pode ser usado de forma on-line no site Dali Server (através do link http://ekhidna2.biocenter.helsinki.fi/dali/), onde pode ser feita a comparação de estruturas usando três bancos de dados: PDB, PDB25 e AlphaFold (AF-DB) [7].

O Dali Server oferece comparações personalizadas como: (i) comparação de uma proteína contra todas as outras do *dataset*, (ii) comparação de todas as proteínas entre si, (iii) comparação de uma proteína com um conjunto de até outras dez proteínas. Com isso, é possível obter uma lista de proteínas similares com dados referentes à semelhança entre as proteínas. Alguns desses dados são o Z-score, o RMSD e Nres.

O Z-score é uma métrica que indica o grau de semelhança entre as proteínas, de modo que, quanto maior o seu valor, mais similares as proteínas são [8]. Essa métrica é definida como:

$$Z_{QT} = \frac{\mathrm{DALI}_{QT} - m(L)}{\sigma(L)} \quad \text{(c)}$$

onde L é a média geométrica do tamanho das estruturas Q e T, dada pela expressão

$$L = \sqrt{L_Q L_T} \quad \text{(d)}$$

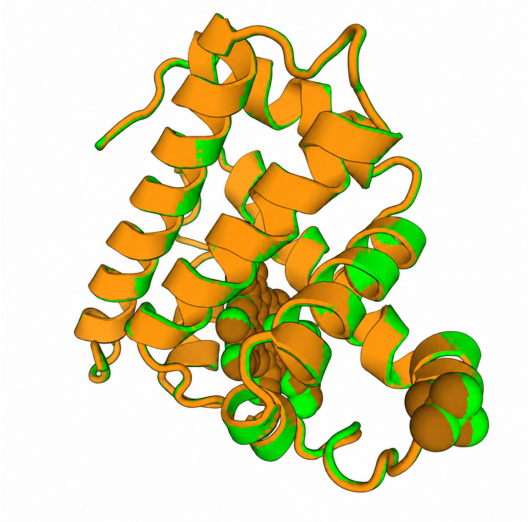
e os parâmetros m e σ são o *score* médio e o desvio padrão, respectivamente.

O RMSD é a métrica usada para indicar a diferença média entre os átomos de duas estruturas de proteínas alinhadas [5]. Essa métrica é dada por:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i^Q - r_i^T\right)^2} \quad \text{(e)}$$

onde N é o número de átomos, C_α o carbono alfa e r_i^Q e r_i^T são as coordenadas vetoriais das proteínas Q e T, respectivamente. Essa é uma medida que relaciona as estruturas atômicas uma com a outra após o alinhamento das proteínas, de modo que, quanto mais próxima de zero, mais parecidas as estruturas são. Já o Nres (número de resíduos alinhados) mostra quantos aminoácidos das proteínas foram pareados.

Figura 2. Sobreposição das proteínas 101M-A e 1DTM-A



Fonte: Elaborado pelos autores (2026).

O DALI constrói uma matriz de distância para cada proteína, onde cada elemento representa a distância entre um par de átomos C_α. Em seguida, as matrizes de distância das duas proteínas são comparadas e é feita busca em regiões com padrões de distância semelhantes, de forma que os resíduos correspon-

dentes às regiões com padrões de distância semelhantes sejam alinhados. Essa abordagem baseada em distâncias permite identificar semelhanças estruturais mesmo em proteínas com diferentes topologias ou com grandes desvios locais. Esse algoritmo permite identificar relações evolutivas distantes através da detecção de homologia estrutural (semelhança tridimensional) em proteínas com baixa similaridade de sequência. Também é possível fazer a análise de interfaces proteicas, ou seja, a análise das superfícies de contato das proteínas que interagem entre si. Isso permite o estudo das interações entre proteínas e outras moléculas.

O exemplo da Figura 2 traz o alinhamento par a par das proteínas teste 101M-A e 1DTM-A. Após o alinhamento, foi possível observar que, dos 154 resíduos da proteína 101M-A e dos 152 resíduos da proteína 1DTM-A, 152 resíduos foram alinhados. Isso significa que essas proteínas são expressivamente similares. O RMSD foi de 0,5, o que sugere que as duas proteínas têm estruturas altamente similares. O Z-score foi de 28,1, indicando alto grau de semelhança. Essas informações comprovam a sobreposição das proteínas mostradas na Figura 2.

2.3 Proteínas Utilizadas

As proteínas analisadas foram obtidas do banco de dados Protein Data Bank (PDB), disponível em https://www.rcsb.org/?ref=nav_home, repositório amplamente utilizado na bioinformática estrutural. Além disso, como será abordado na Seção 3, cada uma dessas proteínas é formada por uma ou mais cadeias polipeptídicas que apresentam uma nomenclatura própria no PDB e uma equivalente no DALI. Nos resultados a seguir, iremos adotar a nomenclatura do DALI para cada uma dessas estruturas que apresenta a forma “s + N + X”, onde N é um número natural que corresponde ao valor da estrutura na comparação entre proteínas (1 para a primeira proteína, 2 para a segunda, etc.) e X corresponde à cadeia equivalente do PDB (como A, B, entre outras).

As proteínas 1M17, 1QRA, 2R6G e 3V2A, todas associadas a processos oncogênicos, foram comparadas neste estudo:

- A proteína 1M17 (formada pela cadeia A de acordo com o PDB) está relacionada ao receptor do fator de crescimento epidérmico (EGFR), que controla o crescimento e a divisão das células. Quando esse receptor funciona de forma anormal, pode estimular o crescimento descontrolado de células, estando associado a diversos tipos de câncer.
- A 1QRA (tem cadeia A) representa uma proteína da família Ras, responsável por transmitir sinais dentro da célula. Ela atua como um “interruptor” que liga ou desliga processos de crescimento celular. Alterações nessa proteína são comuns em cânceres, pois podem manter esse sinal sempre ativo.
- A proteína 3V2A (que apresenta as cadeias A e R) está ligada ao receptor do fator de crescimento do endotélio vascular (VEGFR), que participa da formação de novos vasos sanguíneos (angiogênese). Esse processo é importante para tumores, pois permite que eles recebam nutrientes e cresçam, além de facilitar a disseminação do câncer.
- A 2R6G (formada pelas cadeias A, B, E, F e G) é uma proteína envolvida no transporte de moléculas utilizando energia (ATP), encontrada principalmente em bactérias. Embora não esteja diretamente ligada ao câncer humano, proteínas semelhantes podem influenciar a resistência de células tumorais a medicamentos.

3 RESULTADOS

Os algoritmos TM-Align e o DALI foram utilizados para realizar o alinhamento estrutural de proteínas associadas ao câncer, com o objetivo de analisar similaridades e diferenças estruturais. As proteínas 1M17, 1QRA, 2R6G e 3V2A, todas associadas a processos oncogênicos, foram alinhadas duas a duas. Todos os alinhamentos foram realizados através das plataformas *on-line* Zhang Lab (https://aideepmed.com/TM-align/) e Dali Server (http://ekhidna2.biocenter.helsinki.fi/dali/).

Para cada par de proteínas, foram realizadas comparações utilizando ambos os algoritmos, possibilitando uma avaliação abrangente das relações estruturais. Os dados obtidos foram analisados por meio das métricas TM-score, RMSD, resíduos de cada proteína e L_{ali} (resíduos alinhados), fornecidas pelo TM-Align, e pela métrica Z-score, fornecida pelo DALI.

3.1 Comparação par a par

A Tabela 1 mostra o resultado da comparação par a par entre as proteínas usando o TM-Align. Em cada linha estão dispostas as informações obtidas para cada par. Os dois valores do TM-score por par se devem à normalização feita no cálculo dessa métrica com base no tamanho de cada proteína (conforme indicado na Seção 2.1). Por consequência, proteínas de tamanhos diferentes apresentam *scores* diferentes, porém que seguem a mesma lógica: *scores* próximos a zero indicam baixa similaridade. O RMSD indica alta dispersão entre os átomos das proteínas alinhadas, o que juntamente com o L_{ali} mostra que essas proteínas não apresentam estruturas semelhantes.

Tabela 1. Comparação entre os pares de proteínas utilizando TM-Align

Par de proteínas	TM-score	RMSD	Resíduos	L_{ali}
1M17 / 1QRA	0,22 / 0,35	5,41	312 / 166	103
1M17 / 2R6G	0,26 / 0,23	6,67	312 / 371	144
1M17 / 3V2A	0,14 / 0,20	5,86	312 / 185	69
1QRA / 2R6G	0,38 / 0,21	5,35	166 / 371	114
1QRA / 3V2A	0,22 / 0,20	4,31	166 / 185	58
2R6G / 3V2A	0,21 / 0,34	5,96	371 / 185	120

Fonte: Elaborado pelos autores (2026).

Conforme a Tabela 1, o par 1M17 / 1QRA apresentou TM-scores de 0,22 e 0,35, com RMSD igual a 5,41 Å. Apesar de ambas participarem de vias de transdução de sinal, o alinhamento estrutural foi limitado a apenas 103 resíduos, indicando que as similaridades observadas se restringem a regiões locais da estrutura. A diferença entre os TM-scores ocorre devido à discrepância entre os tamanhos das proteínas, uma vez que 1M17 possui 312 resíduos enquanto 1QRA apenas 166 resíduos.

Na comparação entre 1M17 / 2R6G, os TM-scores de 0,26 e 0,23, juntamente com RMSD de 6,67 Å, evidenciam elevada divergência estrutural. Embora ambas utilizem ATP, seus mecanismos funcionais são distintos, refletindo-se diretamente em suas conformações tridimensionais. O valor relativamente elevado de L_{ali} = 144 indica que uma quantidade considerável de resíduos pôde ser alinhada, porém com baixa qualidade geométrica global.

O par 1M17 / 3V2A apresentou TM-scores ainda menores (0,14 e 0,20), sugerindo uma similaridade estrutural muito baixa, mesmo ambas sendo proteínas da classe das tirosina quinases. O alinhamento envolveu apenas 69 resíduos, indicando que as regiões conservadas provavelmente correspondem apenas a motivos funcionais específicos associados ao domínio quinase.

Na comparação 1QRA / 2R6G, obteve-se o maior TM-score da tabela (0,38), embora ainda abaixo do limiar associado à similaridade estrutural significativa. Esse resultado sugere a existência de algumas regiões estruturais parcialmente compatíveis, possivelmente relacionadas à interação com nucleotídeos trifosfato. Entretanto, o RMSD de 5,35 Å indica que o alinhamento global permanece bastante disperso.

O par 1QRA / 3V2A apresentou TM-scores de 0,22 e 0,20 e o menor RMSD observado entre os pares analisados (4,31 Å). Apesar disso, o alinhamento envolveu apenas 58 resíduos, indicando que a menor dispersão ocorre em uma região estrutural reduzida e não representa similaridade global entre as proteínas.

Por fim, a comparação entre 2R6G / 3V2A apresentou TM-scores de 0,21 e 0,34, com RMSD de 5,96 Å e alinhamento de 120 resíduos. Embora ambas utilizem ATP em seus mecanismos biológicos, a organização estrutural das proteínas permanece bastante distinta, especialmente devido à natureza multimérica de 2R6G e à organização típica de receptores tirosina quinase em 3V2A.

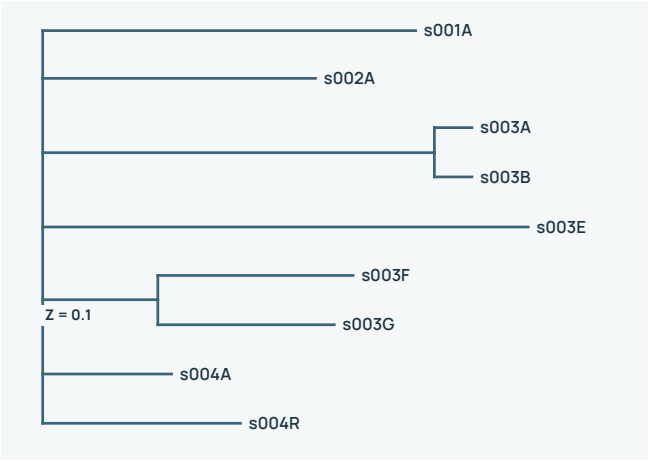
De maneira geral, os resultados obtidos pelo TM-Align indicam que os pares analisados possuem apenas similaridades estruturais locais, insuficientes para caracterizar dobramentos tridimensionais globalmente conservados. Os baixos valores de TM-score, associados aos elevados valores de RMSD, mostram que as proteínas apresentam conformações expressivamente diferentes, ainda que compartilhem alguns mecanismos bioquímicos ou funções celulares relacionadas.

Também foram feitas comparações duas a duas utilizando o algoritmo DALI seguindo a mesma ordem da primeira coluna da Tabela 1. Os resultados foram obtidos por meio da opção de comparação *All against all* da plataforma e, posteriormente, selecionando pares de proteínas. Em todas as comparações, obteve-se um Z-score de 0,1, indicando uma similaridade estrutural insignificante entre essas proteínas, embora possam apresentar alguma semelhança em funções estruturais.

3.2 Comparação complementar

Não foi possível utilizar a opção *Pairwise* para obter as demais métricas do DALI (RMSD e Nres), pois, com o baixíssimo Z-score, o algoritmo não reportou resultados detalhados, conforme esperado. Então, com o intuito de complementar a análise, foi feita uma comparação entre as quatro proteínas simultaneamente através da mesma opção *All against all* desse algoritmo. Foi gerado um dendrograma para representar a relação de similaridade estrutural entre todas as cadeias proteicas analisadas simultaneamente no DALI e com isso identificar agrupamentos locais entre algumas cadeias específicas, conforme ilustrado na Figura 3.

Figura 3. Dendrograma da comparação das quatro proteínas



Fonte: Elaborado pelos autores (2026).

Nessa figura a cadeia s001A, correspondente à proteína 1M17, encontra-se isolada em um ramo superior do dendrograma, indicando baixa proximidade estrutural em relação às demais proteínas analisadas. Esse comportamento é coerente com as comparações individuais envolvendo 1M17, nas quais apenas pequenas regiões alinháveis foram identificadas.

A cadeia s002A, associada à proteína 1QRA, também aparece relativamente isolada, reforçando a ideia de que essa proteína possui organização estrutural distinta das demais. Apesar disso, sua posição intermediária sugere a existência de alguns elementos estruturais parcialmente compartilhados com outras proteínas envolvidas em transdução de sinal.

As cadeias s003A e s003B, pertencentes à proteína 2R6G, formam o agrupamento mais próximo observado no dendrograma. Isso indica elevada similaridade estrutural entre essas cadeias, o que é esperado em proteínas multiméricas (com várias cadeias). De forma semelhante, as cadeias s003F e s003G também aparecem agrupadas em um mesmo ramo, sugerindo repetição de padrões estruturais dentro da própria proteína 2R6G. A cadeia s003E, também pertencente à 2R6G, aparece mais distante das cadeias anteriores, indicando que algumas subunidades dessa proteína podem desempenhar funções estruturais ou conformacionais distintas dentro da mesma proteína.

Por fim, as cadeias s004A e s004R, associadas à proteína 3V2A, aparecem separadas entre si e relativamente afastadas das demais cadeias. Esse comportamento sugere heterogeneidade estrutural interna na proteína 3V2A, compatível com as análises par a par anteriores, nas quais diferentes cadeias apresentaram graus distintos de similaridade estrutural.

De maneira geral, o dendrograma evidencia que as maiores similaridades estruturais ocorrem entre cadeias pertencentes à mesma proteína, especialmente no caso de 2R6G, enquanto as comparações entre proteínas distintas resultam em baixa similaridade global. Isso reforça que, embora algumas proteínas compartilhem funções biológicas relacionadas, como interação com ATP ou participação em vias de sinalização celular, suas arquiteturas tridimensionais permanecem expressivamente diferentes.

4 DISCUSSÃO

Os resultados obtidos mostram que os algoritmos DALI e TM-Align possuem características complementares na análise estrutural de proteínas. Embora ambos tenham indicado baixa similaridade estrutural global entre as proteínas analisadas, cada método contribuiu de forma distinta para a interpretação dos alinhamentos. O TM-Align destacou-se na avaliação quantitativa da sobreposição estrutural, utilizando métricas como TM-score e RMSD para medir similaridade e dispersão geométrica. Mesmo com TM-scores inferiores a 0,5, o algoritmo identificou alinhamentos locais relevantes, sugerindo a presença de motivos estruturais parcialmente conservados relacionados à interação com ATP.

Por outro lado, o DALI apresentou maior sensibilidade na análise qualitativa das relações estruturais. A análise *All against all* permitiu identificar agrupamentos locais importantes, especialmente entre cadeias da proteína 2R6G, evidenciando padrões estruturais recorrentes em proteínas multiméricas. Além disso, o método mostrou que, apesar de algumas semelhanças funcionais, as proteínas possuem arquiteturas tridimensionais bastante distintas.

No contexto da oncologia estrutural, essa complementaridade é especialmente relevante, pois proteínas associadas ao câncer frequentemente apresentam alterações localizadas sem modificar toda a estrutura global. Assim, enquanto o TM-Align auxilia na avaliação quantitativa de mudanças conformacionais, o DALI contribui para a identificação de motivos estruturais conservados e relações estruturais mais sutis. Dessa forma, o uso combinado dos algoritmos proporciona uma análise estrutural mais abrangente e biologicamente informativa.

5 CONCLUSÃO

Os resultados obtidos demonstram que os algoritmos DALI e TM-Align possuem características complementares fundamentais para a análise estrutural de proteínas associadas ao câncer. Enquanto o DALI apresenta maior sensibilidade na identificação de homologias distantes, regiões funcionais conservadas e relações estruturais sutis, o TM-Align se destaca pela precisão quantitativa e eficiência na sobreposição estrutural global por meio do TM-score. Dessa forma, os dois métodos fornecem perspectivas distintas, porém complementares, sobre a organização tridimensional das proteínas.

No contexto da oncologia estrutural, essa complementaridade é particularmente relevante, pois muitas proteínas relacionadas ao câncer apresentam mutações localizadas que alteram regiões funcionais específicas sem necessariamente modificar toda a estrutura global. Nesse cenário, o uso combinado dos algoritmos permite tanto avaliar mudanças conformacionais globais quanto identificar motivos estruturais conservados associados a funções biológicas importantes e possíveis relações evolutivas.

Assim, este trabalho evidencia que abordagens integradas de alinhamento estrutural contribuem para uma compreensão mais abrangente dos mecanismos moleculares envolvidos na oncogênese, auxiliando na interpretação funcional de alterações estruturais e no potencial desenvolvimento de novos alvos terapêuticos.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Agradecimentos. Os autores agradecem ao Laboratório de Computação de Alto Desempenho (LabCAD) pelo apoio na realização dos experimentos, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por meio do Código de Financiamento 001, e à Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), pelo apoio financeiro concedido por meio do Edital APQ1 nº 26/2021 (Processo nº E-26/211.798/2021).

Financiamento. Os autores não declararam fonte de financiamento.

Disponibilidade de dados e materiais. Os autores não declararam a disponibilidade dos dados e materiais da pesquisa.

Conflitos de interesse. Os autores não declararam conflitos de interesse.

Declaração de uso de IA. Declaramos que, durante a elaboração deste trabalho, foi utilizada a ferramenta de inteligência artificial ChatGPT, desenvolvida pela OpenAI, exclusivamente como apoio na revisão linguística e aprimoramento da redação do texto acadêmico. O uso da ferramenta restringiu-se à correção gramatical, sugestão de melhorias de clareza, coesão e organização textual, não tendo sido empregada para a geração de conteúdo científico original, formulação de resultados, análises ou conclusões. Ressalta-se que todo o conteúdo técnico-científico, a interpretação dos resultados, a estrutura do trabalho e as conclusões apresentadas são de inteira responsabilidade dos autores.

REFERÊNCIAS

1. Baker D, Sali A. Protein structure prediction and structural genomics. Science. 2001;294(5540):93-6. doi: https://doi.org/10.1126/science.1065659
2. Holm L. Using Dali for Protein Structure Comparison. In: Methods in Molecular Biology. New York, NY: Springer US; 2020. p. 29-42. doi: https://doi.org/10.1007/978-1-0716-0270-6_3
3. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol. 1993;233(1):123-38. doi: https://doi.org/10.1006/jmbi.1993.1489
4. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33(7):2302-9. doi: https://doi.org/10.1093/nar/gki524
5. Hasegawa H, Holm L. Advances and pitfalls of protein structural alignment. Curr Opin Struct Biol. 2009;19(3):341-8. doi: https://doi.org/10.1016/j.sbi.2009.04.003
6. Collier JH, Allison L, Lesk AM, Garcia de la Banda M, Konagurthu AS. A new statistical framework to assess structural alignment quality using information compression. Bioinformatics. 2014;30(17):i512-8. doi: https://doi.org/10.1093/bioinformatics/btu460
7. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res. 2022;50(W1):W210-W215. doi: https://doi.org/10.1093/nar/gkac387
8. Holm L. DALI and the persistence of protein shape. Protein Sci. 2020;29(1):128-40. doi: https://doi.org/10.1002/pro.3749

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde