

## Caracterização estrutural de proteínas por métricas de alta e baixa resolução

**Vanessa Cadan Scheffer<sup>1</sup>; Marcelo Augusto dos Reis<sup>2</sup>**

### RESUMO

A função biológica de uma proteína está intrinsecamente associada à forma tridimensional assumida por este tipo de macromolécula biológica. Motivados pelo interesse em conhecer as funções desempenhadas por tais moléculas, pesquisadores da área de biofísica e bioinformática têm investido no desenvolvimento de técnicas para a predição estrutural de proteínas.

Diante deste contexto, realizamos um estudo comparativo com resultados obtidos através de simulações da técnica experimental de espalhamento de raios X a baixos ângulos, conhecida como SAXS, e com resultados obtidos pelo *software* TM-score, no qual a métrica é feita pela comparação direta proteína-proteína.

O trabalho foi desenvolvido a partir de um viés computacional, com algoritmos que quantificam a qualidade da combinação entre estruturas.

**Palavras-chave:** Bioinformática, proteínas, SAXS, modelagem computacional.

### INTRODUÇÃO

“Proteínas são macromoléculas biológicas que desempenham funções vitais em um organismo” (LEHNINGER, 2008), desde a respiração celular ao combate de doenças. Estudos de biofísica molecular vêm confirmando a ideia de a atividade estar intrinsecamente relacionada com a conformação tridimensional assumida por uma proteína. Portanto, o estudo da estrutura tridimensional de uma proteína é essencial para se conhecer os possíveis mecanismos de ação que viabilizam a especificidade da função proteica.

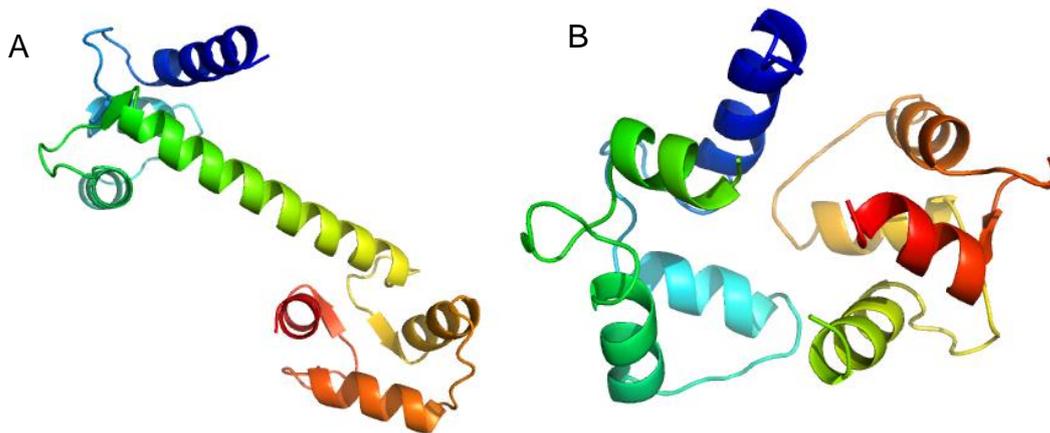
Um grupo seletivo de proteínas exibe características estruturais conhecidas como estado aberto e fechado (Figura 1), esses termos são usados para descrever

---

<sup>1</sup> Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Câmpus Inconfidentes. Inconfidentes/MG, email: 10022025@ifs.ifsuldeminas.edu.br;

<sup>2</sup> Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Câmpus Inconfidentes. Inconfidentes/MG, email: marcelo.reis@ifsuldeminas.edu.br;

o estado relaxado e contraído de uma proteína. As reações químicas que promovem esses estados não são objetos de estudo nesse trabalho. A análise se dará exclusivamente do ponto de vista estrutural. “Movimentos de grande amplitude podem estar relacionados às transições conformacionais do tipo estado aberto ↔ estado fechado, sabidamente relacionadas à função proteica”. (TAMA, 2001, p.1)



**Figura 1:** (A) Proteína Calmodulina (PDB ID: 1c1l) no estado aberto; (B) Proteína Calmodulina (PDB ID: 1ctr) no estado fechado.

Existem diversos métodos já sendo utilizados, e outros em desenvolvimento, para predição de estruturas, cada um com sua especificidade técnica, tanto no que diz respeito à mão de obra especializada, quanto aos recursos computacionais. A técnica de SAXS permite conhecer formas oligoméricas e distinguir arranjos tridimensionais. É uma técnica experimental, na qual, uma proteína em solução é submetida aos raios X. Os elétrons dos átomos interagem com a radiação incidente espalhando-a. Disso decorre um perfil relacionado à intensidade do espalhamento em função do ângulo permitindo uma análise das características geométricas globais da estrutura proteica. Contudo, é possível simular os dados de SAXS (função  $I(q)$ ) permitindo aferir suas potencialidades e limitações com a vantagem de utilização de recursos mínimos de pesquisa.

Uma técnica computacional também utilizada para quantificar similaridades estruturais é o TM-score (ZHANG; SKOLNICK, 2004). O algoritmo relacionado à métrica trabalha com informação mais precisa caracterizando-a como uma técnica de alta resolução. As proteínas são alinhadas de uma maneira ótima, e então é calculada a distância entre os pares de resíduos correspondentes.

Nosso objetivo é verificar o grau de correlação entre as métricas obtidas pelas duas técnicas. Em geral, não se conhece a estrutura de uma proteína no início de um estudo inédito, no qual apenas é possível se obter o valor conhecido como  $\chi$  “CHI”, que representa uma comparação quantitativa entre as curvas de intensidade  $I(q)$  de um modelo proteico e a curva de intensidade experimental (Equação 2). Portanto faz-se necessário compreender até que ponto o resultado gerado pelo métrica baseada em SAXS se correlaciona com uma métrica mais acurada como o TM-score.

## MATERIAL E MÉTODOS

O uso de *software* livre foi largamente empregado para que não seja necessária a aquisição de licenças. Podem ser destacados alguns programas mais utilizados neste trabalho, como por exemplo: o sistema operacional LINUX (Ubuntu), visualizador de estruturas tridimensionais PyMOL (<http://www.pymol.org/>), editor de gráficos GNUPLOT (<http://www.gnuplot.info/>), programa que simula os dados de SAXS (módulo SAXSSCORE *do pipeline* SAXSTER (REIS; APARICIO; ZHANG, 2011) e a métrica estrutural TM-score (<http://zhanglab.ccmb.med.umich.edu/TM-score/>).

A partir do banco de dados PDB (<http://www.rcsb.org/>), 36 proteínas - 18 pares de estado aberto e 18 fechado - com estruturas cristalográficas abertas e fechadas foram coletadas a partir da referência disponível na literatura TAMA e SANEJOUAND (2001, p. 2).

A utilização dos algoritmos: TM-score (Equação 1), e SAXSSCORE, que gera o valor  $\chi$  conhecido por “CHI” (Equação 2) foi empregada tendo em vista as seguintes métricas:

$$\text{TM-score} = \text{Max} \left[ \frac{1}{L_N} \sum_{i=1}^{L_T} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2} \right] \quad (1)$$

O resultado do TM-score está compreendido entre [0,1] onde:

- 1 representa a sobreposição perfeita;
- Valores > 0.5 representam uma sobreposição satisfatória;

- Valores < 0.17 significam que as proteínas comparadas não pertencem a mesma família.

$$\chi = \frac{1}{n} \sum_{k=1}^n \left\{ \frac{I_{exp}(q_k) - cICG(q_k)}{\sigma(q_k)} \right\} \quad (2)$$

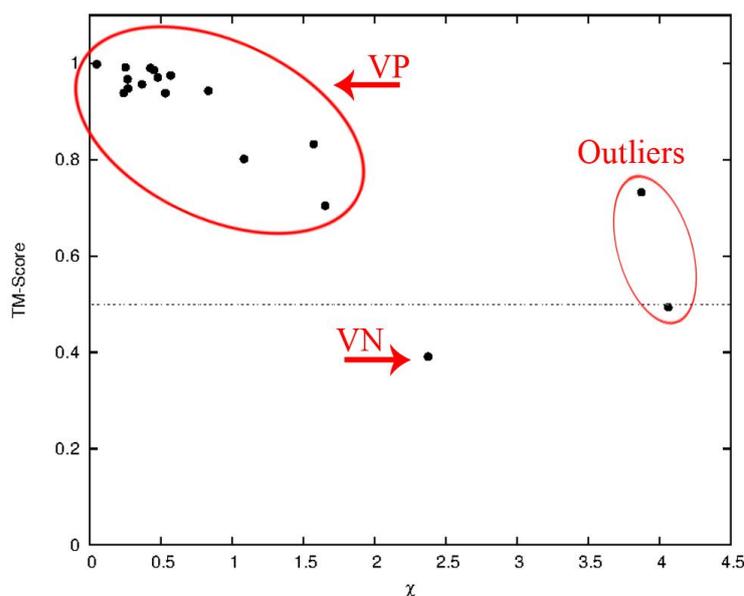
Em seguida os resultados foram correlacionados utilizando o coeficiente de correlação de Pearson, dado pela equação:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

## RESULTADOS E DISCUSSÃO

Após utilizar as técnicas citadas acima para quantificar a sobreposição das proteínas no estado aberto e fechado, concluímos que o TM-score e o SAXSCORE apresentam uma razoável correlação, embora *outliers* sempre existam, pois, o TM-score analisa e pontua sobre domínios enquanto o SAXSCORE faz uma análise global da estrutura, resultado da baixa resolução da informação obtida pelo experimento de SAXS. O resultado obtido através do coeficiente de correlação de Pearson foi de -0.71.

Diante essas características separamos os resultados em três grupos como mostra a Figura 2:



**Figura 2:** Gráfico que relaciona os resultados obtidos pelo SAXSCORE (eixo x), e pelo TM-score (eixo y). Os 18 resultados foram agrupados em 3 categorias: (VP) Verdadeiro positivo, para os resultados igualmente positivos; (VN) Verdadeiro negativo, para os resultados igualmente negativos; Outliers são os pontos que se afastam da correlação desejada para a relação de TM-score e  $\chi$ .

Das 18 proteínas analisadas, 15 pertencem ao grupo verdadeiro positivo VP, no qual o TM-score próximo à unidade indica que os domínios comparados entre o estado aberto e fechado possuem uma grande similaridade, enquanto o  $\chi$  próximo à zero representa que o perfil gerado por SAXS encontrou uma grande similaridade na visão global da proteína. Estruturas otimamente sobrepostas e com estruturas muito similares elevam o coeficiente de relação das duas medidas.

Para o caso de verdadeiro negativo VN, o TM-score apresenta valores mais próximos de zero, sendo assim, as estruturas comparadas possuem significativa diferença em sua estrutura, pois TM-score trabalha com informações precisas de alta resolução. Analisando o resultado de  $\chi$ , quanto mais distante de zero o valor está, mais diferenças existem entre as estruturas, porém ainda não existe uma quantificação que relacione exatamente o valor com o grau de diferença das estruturas resultantes de SAXS. Nesse caso o verdadeiro negativo foi confirmado devido à comparação entre as duas respostas TM-score e SAXSCORE.

Nos dois casos de *outliers*, é importante observar que ambos estão fora dos valores esperados para a comparação. O desejável seria existir uma correlação entre TM-score e SAXSCORE, isto é, quando TM-score  $\sim 1$ , o segundo tenderá a valores próximos à zero, e quando o TM-score estiver próximo à zero, o SAXSCORE fique cada vez mais alto. Esses pontos de *outliers* acarretam a diminuição do resultado do coeficiente de correlação de Pearson.

## CONCLUSÕES

O uso do espalhamento de raios X a baixos ângulos (SAXS) para predição de estruturas de proteínas apresenta resultados satisfatórios para uma análise global de uma molécula de proteína. Isto foi evidenciado através do coeficiente de Pearson da ordem de -0.71 entre as métricas TM-score e  $\chi$ . Entretanto, devido às características inerentes ao SAXS, a curva que é utilizada para a análise é uma informação de baixa resolução, o que pode ocasionar falhas no reconhecimento de

um modelo de proteína, o que explica a ocorrência de *outliers*. Este resultado reforça a idéia do contínuo desenvolvimento de técnicas computacionais que combinem dados experimentais com dados de alta resolução para uma melhor acurácia do método preditivo.

## REFERÊNCIAS BIBLIOGRÁFICAS

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. W. H. Freeman, fifth edition ed., 2008. 1152 p.

TAMA, F.; SANEJOUAND, Y.-H. Conformational change of proteins arising from normal mode calculations. **Protein Engineering**, vol.14, p. 1 - 6, 2001.

TIRION, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. **The American Physical Society**, vol. 77, p. 1905 - 1908, 1996.

ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins**, vol. 57, p. 702 - 710, 2004.

REIS, M. A. dos; APARICIO, R.; ZHANG, Y. Improving protein template recognition by using small angle X-ray scattering profiles. **Biophysical Journal**, v. 101, p. 2770-2781, 2011.

## AGRADECIMENTOS



Agradecemos ao IFSULDEMINAS – Câmpus Inconfidentes, pelo apoio financeiro através de bolsa de iniciação científica via EDITAL N°05/2013 para fomento interno de projeto de pesquisa.