

Uma modelagem de estrutura de proteína utilizando ferramentas de bioinformática

Alex D. Camargo¹, Érico M. H. Amaral¹, Sandro S. Camargo¹

¹Universidade Federal do Pampa (UNIPAMPA)
Bagé – RS – Brasil

{alexcamargo, ericoamaral, sandrocamargo} @unipampa.edu.br

Abstract. *The advent of computational technologies, integrated with knowledge in molecular biology, has permitted projection and simulation of protein structures in third dimension. Hereditary deficiency of the lysosomal enzyme B-galactosidase (coded by GLB1) leads to two clinically distinct human diseases: the GM1 Gangliosidosis and the Morquio B. Aiming understand the effects of 3THC protein structure coded by GLB1, its modeling was realized through bioinformatics tools: Modeller and the Phyre2.*

1. Introdução

Podemos considerar a Bioinformática como uma linha de pesquisa que envolve aspectos multidisciplinares e que surgiu a partir do momento em que se iniciou a utilização de ferramentas computacionais para a análise de dados genéticos, bioquímicos e de biologia molecular [Prosdocimi 2011]. A presente pesquisa trata das doenças humanas, clinicamente distintas, causadas pela deficiência hereditária da enzima lisossômica B-galactosidase codificada por GLB1, a Gangliosidose GM1 e Morquio B. Os pacientes com Gangliosidose GM1 mostram graus variados de neurodegeneração e anormalidades esqueléticas, já os pacientes com Morquio B apresentam displasia esquelética e opacidade de córnea, sem envolvimento do sistema nervoso central [Sperb 2012]. As principais técnicas de determinação experimental de estruturas tridimensionais, como a cristalografia por difração de raios-x e ressonância magnética nuclear, são consideradas financeiramente custosas e demoradas. Atualmente, existem diversas estratégias e métodos disponíveis para a construção de modelos de proteínas por modelagem comparativa *in-silico* (através do computador). O principal objetivo desta pesquisa é prever a estrutura tridimensional da proteína 3THC, codificada pelo gene GLB1, tendo em vista que o modelo disponível no *Protein Data Bank*, atualmente a maior fonte pública de estruturas de proteínas, não tem determinado os aminoácidos 300, 301, 302, 440 e 448.

2. Solução proposta

Para este trabalho, realizou-se a predição da proteína alvo através das ferramentas computacionais: *Modeller* (com a interface *EasyModeller*) e *Phyre2*. O *Modeller*¹ é um robusto preditivo programa de modelagem comparativa. Tal aplicação é capaz de construir modelos de proteínas com base na sequência de aminoácidos da proteína alvo e dados estruturais de uma ou mais estruturas 3D da proteína molde, preferencialmente homólogas à proteína alvo, com identidade sequencial superior a 25% [Machado 2013]. Nesta pesquisa também foi utilizado o *Phyre2*, classificado entre os melhores programas de predição de estrutura terciária, segundo dados do CASP (*Critical Assessment of protein Structure Prediction*) de 2010, sendo amplamente utilizado pela comunidade biológica [Kelley e Sternberg 2009].

¹ Disponível para *download* em:
https://salilab.org/modeller/download_installation.html

3. Considerações parciais

Definida a metodologia para a condução do experimento e as configurações necessárias nas ferramentas, passou-se então aos testes propriamente ditos. O estudo incluiu vários problemas e métodos relacionados na bibliografia, porém, somente alguns foram implementados em função da grandeza de soluções. Para a predição no *Modeller*, a sequência alvo 3THC foi submetida ao algoritmo BLAST (*Basic Local Alignment Search Tool*) [Altschul 1997], através do *website* NCBI (*National Center for Biotechnology Information*) para que sejam realizadas buscas por proteínas moldes através do alinhamento. Entre as interfaces que estão disponíveis e atualizadas, optou-se pela facilidade de instalação e operação, o *EasyModeller*. Já o *Phyre2*, serviço *online*, apresentou grande vantagem pois basta a submissão da sequência alvo (formato FASTA) para que a predição seja realizada. Ao final do processo, é possível visualizar várias informações sobre o modelo tridimensional, em contrapartida não podemos saber com detalhes quais procedimentos foram adotados pela ferramenta. A construção do modelo (formato PDB) gerado através do *Modeller* e pelo *Phyre2* está ilustrado nas Figuras 1 e 2. A etapa subsequente do trabalho envolverá a avaliação e refinação dos modelos obtidos pelo computador, o que favorecerá a compreensão dos efeitos causados pela proteína estudada.

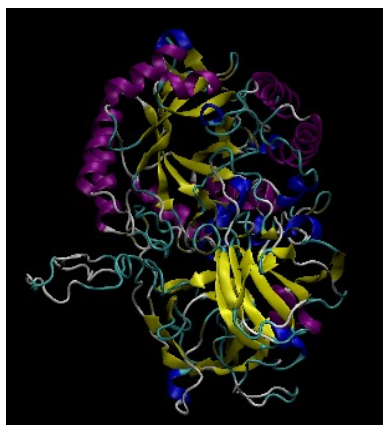


Figura 1. Modelo tridimensional: *Modeller*

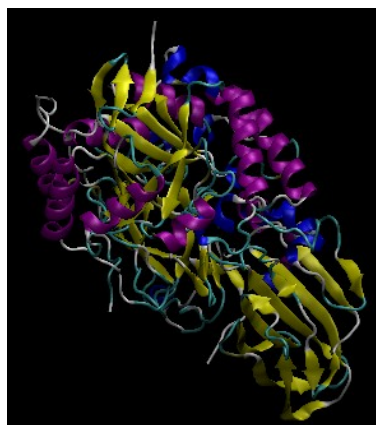


Figura 2. Modelo tridimensional: *Phyre2*

Referências

- Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Miller, W.; Lipman, D. J. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". *Nucleic Acids Res.* 25:3389-3402, 1997.
- Kelley, L. A.; Sternberg, M. J. "Protein structure prediction on the web: a case study using the Phyre server *Nature Protocols*". Nature Publishing Group, 2009.
- Machado, A. L. G. "Descoberta e Planejamento de Novos Agroquímicos para a Cultura da Cana-de-Açúcar". Dissertação de Mestrado. Programa de Pós-Graduação em Física Aplicada Biomolecular, 2013.
- Prosdoci, F. "Introdução à Bioinformática". Instituto de Química, 2007. Disponível em: <<http://www.iq.usp.br/setubal/bmc/2013/>>. Acesso em: 05/10/2014.
- Sperb, L. F. "Gangliosidose GM1: aspectos clínicos e moleculares da população brasileira e a busca de novas terapias para o tratamento da doença". Tese de Doutorado. Programa de Pós-graduação em Genética e Biologia Molecular, 2012.