

Desenvolvimento e implementação de rotinas de modelagem molecular por homologia e *ab initio* aplicadas às proteínas multidomínio.

Centro Nacional de Pesquisa em Energia e Materiais (CNPEM)
Laboratório Nacional de Biociências (LNBio)
Laboratório de Biologia Computacional (LBC)

Pesquisador responsável: Rodrigo Honorato (LNBio)

Pré-requisitos: Espera-se do candidato habilidades de leitura e compreensão de publicações acadêmicas em língua inglesa, além de conhecimentos básicos de biologia estrutural. O candidato selecionado deverá demonstrar familiaridade com ambiente Linux e desenvolvimento em Python. A seleção é aberta a alunos de ciências biológicas e exatas.

1 Introdução

1.1 Proteínas Multidomínio

Um domínio proteico é uma sequência de aminoácidos evolutivamente conservada e funcionalmente independente. Uma determinada proteína pode ser composta por mais de um domínio, caracterizando-a como uma proteína multidomínio.

O agrupamento de múltiplos domínios em uma mesma proteína cria uma entidade funcional distinta, que tem como função a combinação de cada domínio. As variações dessas combinações são o mecanismo mais importante para a evolução funcional nos genomas eucarióticos. O entendimento de como as proteínas multidomínio ganham ou perdem função é chave para o entendimento da evolução proteica, uma vez que proteínas multidomínio podem ter diversas funções acumuladas. Existe uma diferença quanto a prevalência de proteínas multidomínio em diferentes ramos evolutivos, sendo que maior quantidade desse tipo de proteína é observada em organismos eucariotos, chegando a 70% do proteoma. (HAN et al., 2007).

Sistemas compostos por esse tipo de proteína e que apresentam dobramento independente dos conectores são prevalentes em processos de transcrição, replicação, sinalização e adesão celular (BLACKLEDGE, 2010).

1.2 Implementação **MAD**

Foi desenvolvido no Laboratório de Biologia Computacional (LBC) do LNBio o **MAD** (*Multidomain protein Assembly and Design protocol*) que é capaz de obter modelos estruturais tridimensionais de proteínas multidomínio e prever sua variação conformacional, contribuindo para um melhor entendimento da função e atividade destas proteínas (HONORATO; OLIVEIRA, 2014).

O software conta com uma série de módulos, entre eles duas etapas de modelagem, uma por homologia e uma *ab initio*. A modelagem por homologia, também chamada de modelagem comparativa, tem como objetivo criar modelos para cada um dos domínios contidos na proteína alvo. Atualmente essa etapa é executada através de um software proprietário. Aplica-se modelagem *ab initio* na criação dos conectores flexíveis que serão utilizados nas etapas subsequentes do **MAD**.

2 Objetivos

(1) Desenvolver e automatizar etapa de modelagem por homologia para que sejam obtidos modelos tridimensionais de cada um dos domínios identificados pelo **MAD**. (2) Investigar e melhorar etapa de modelagem *ab initio*.

3 Métodos

3.1 Modelagem por homologia

O software MODELLER (ESWAR et al., 2001) será utilizado como base para criação dos modelos por homologia. Será necessário também automatizar a produção do alinhamento entre a proteína alvo e proteína molde, além da obtenção dos mesmos a partir do *Protein Data Bank* (PDB). Os modelos finais serão avaliados utilizando ferramentas estatísticas e campos de força específicos para modelagem molecular.

3.2 Modelagem *ab initio*

Os conectores entre domínios proteicos são fundamentais para elucidação das possíveis conformações de uma proteína multidomínio. Utilizamos o PyRosetta (CHAUDHURY; LYSKOV; GRAY, 2010) para obter um conjunto de dobramentos prováveis para cada uma dessas regiões flexíveis. Serão investigadas algumas características das regiões conectoras com o objetivo de melhorar (otimizar) a etapa de modelagem *ab initio*.

Referências

BLACKLEDGE, M. Mapping the conformational mobility of multidomain proteins. *Biophysical journal*, Biophysical Society, v. 98, n. 10, p. 2043–4, maio 2010. ISSN 1542-0086. Citado na página 1.

CHAUDHURY, S.; LYSKOV, S.; GRAY, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, Oxford University Press (OUP), v. 26, n. 5, p. 689–691, jan 2010. Citado na página 2.

ESWAR, N. et al. Comparative protein structure modeling using modeller. In: _____. *Current Protocols in Protein Science*. [S.l.]: John Wiley & Sons, Inc., 2001. ISBN 9780471140863. Citado na página 2.

HAN, J.-H. et al. The folding and evolution of multidomain proteins. *Nature reviews. Molecular cell biology*, v. 8, n. 4, p. 319–30, abr. 2007. ISSN 1471-0072. Citado na página 1.

HONORATO, R. V.; OLIVEIRA, P. S. L. Protein-protein docking pipeline using a two-step algorithm with experimental validation. In: *Intelligence Methods for Bioinformatics and Biostatistics*. Cambridge (UK): Department of Informatics, University of Salerno, 2014. p. 112–119. ISBN 9788890643743. Citado na página 2.