

Universidade de São Paulo
Instituto de Física de São Carlos

XIV Semana Integrada do Instituto de
Física de São Carlos

Livro de Resumos da Pós-Graduação

São Carlos
2024

Ficha catalográfica elaborada pelo Serviço de Informação do IFSC

Semana Integrada do Instituto de Física de São Carlos
(13: 21-25 ago.: 2023: São Carlos, SP.)

Livro de resumos da XIII Semana Integrada do Instituto de Física de São Carlos – Universidade de São Paulo / Organizado por Adonai Hilário da Silva [et al.]. São Carlos: IFSC, 2023.
358p.

Texto em português.

1.Física. I. Silva, Adonai Hilário da, org. II. Título.

ISSN: 2965-7679

113

Análise multiparamétrica do problema de fases em cristalografia de proteínas por aprendizado de máquina - caso de estudo: lisozima da clara do ovo de galinha

JUCOVSKI, André Gustavo¹; AMBROSIO, Andre Luis Berteli¹

andre.jucovski@usp.br

¹Instituto de Física de São Carlos - USP

O problema de fases é notório na cristalografia de proteínas por difração de raios-X. A perda experimental de informações sobre as fases das ondas espalhadas construtivamente pelos componentes do cristal se devem às limitações tecnológicas intrínsecas aos sistemas de detecção dessa radiação. (1) Assim, é impossibilitado o cálculo direto da função de distribuição de densidade eletrônica na cela unitária através de uma transformada de Fourier. Atualmente, há dois métodos experimentais que podem ser aplicados para contornar esse problema: (i) a quantificação seletiva do componente dispersivo (λ -dependente) do fator de espalhamento atômico ou (ii) a substituição parcial do solvente aquoso ordenado por íons mais elétron-densos (metálicos ou halogênicos). (2) Alternativamente, informações prévias, na forma de estruturas cristalinas conhecidas que são funcionalmente relacionadas ou homólogas a componentes no cristal, podem servir como fonte de um conjunto inicial de fases. Apesar de desafiadoras, quando viáveis, as aplicações desses diferentes métodos já possibilitaram a determinação de mais de uma centena de milhares de modelos atômicos, para as mais diversas proteínas (e seus complexos). Tendo em vista a existência de uma vasta coleção de informações estruturais já disponibilizadas no banco de dados Protein Data Bank, propõe-se aqui uma análise multiparamétrica do problema das fases, com base em aprendizagem de máquina profunda (AMP). (3) Nossa hipótese é a de que o extensivo mapeamento estatístico de observações sobre distribuições de fases conhecidas, como um modelo preditivo, pode permitir conclusões sobre o valor alvo de fases em conjuntos de dados ainda não resolvidos, eliminando assim a necessidade de experimentos adicionais ou de estruturas homólogas previamente conhecidas. Assim, neste Projeto de Mestrado, estão sendo testados modelos para um seletivo grupo de estruturas de lisozimas provenientes da clara do ovo de galinhas (da sigla HEWL, do inglês Hen egg-white lysozyme), no sentido de avaliar a aplicabilidade e sua reprodutibilidade para outras estruturas no futuro. Até este momento foi obtido êxito em realizar um ciclo completo, isto significa que foi possível a saída da predição das fases por AMP e a utilização destas para a reconstrução dos mapas de densidade eletrônica (MDE). Foi desenvolvido um modelo de predição de forma progressiva que possui uma acurácia média de 15-20 graus, comparando abordagens tradicionais com métricas de Von Mises. Agora tem-se como objetivo refinar ainda mais o modelo e aprimorar a avaliação quantitativa da fidelidade dos MDEs por ARP/wARP.

Palavras-chave: Aprendizado de máquina; Cristalografia; Problema de fases.

Agência de fomento: CAPES (88887.675055/2022-00)

Referências:

- 1 DRENTH, J. **Principles of protein X-ray crystallography**. 3rd ed. New York: Springer, 2007. 332 p.
- 2 RUPP, B. **Biomolecular crystallography: principles, practice, and application to structural biology**. New York: Garland Science, 2010. 809 p.
- 3 GÉRON, A. **Hands-on machine learning with scikit-learn and tensorflow: concepts, tools, and techniques to build intelligent systems**. Sebastopol, CA: O'Reilly Media, 2017. 547 p.