

Universidade de São Paulo  
Instituto de Física de São Carlos

XII Semana Integrada do Instituto de  
Física de São Carlos

Livro de Resumos

São Carlos  
2022

# Semana Integrada do Instituto de Física de São Carlos

SIFSC 12

## Coordenadores

Prof. Dr. Osvaldo Novais de Oliveira Junior

Diretor do Instituto de Física de São Carlos – Universidade de São Paulo

Prof. Dr. Javier Alcides Ellena

Presidente da Comissão de Pós Graduação do Instituto de Física de São Carlos – Universidade de São Paulo

Profa. Dra. Tereza Cristina da Rocha Mendes

Presidente da Comissão de Graduação do Instituto de Física de São Carlos – Universidade de São Paulo

## Comissão Organizadora

Adonai Hilario

Arthur Deponte Zutião

Elisa Goettens

Gabriel dos Santos Araujo Pinto

Henrique Castro Rodrigues

Jefer Santiago Mares

João Victor Pimenta

Julia Martins Simão

Letícia Martinelli

Lorany Vitoria dos Santos Barbosa

Lucas Rafael Oliveira Santos Eugênio

Natasha Mezzacappo

Paulina Ferreira

Vinícius Pereira Pinto

Willian dos Santos Ribela

## Normalização e revisão – SBI/IFSC

Ana Mara Marques da Cunha Prado

Maria Cristina Cavarette Dziabas

Maria Neusa de Aguiar Azevedo

Sabrina di Salvo Mastrantonio

Ficha catalográfica elaborada pelo Serviço de Informação do IFSC

Semana Integrada do Instituto de Física de São Carlos  
(12: 10 out. - 14 out. : 2022: São Carlos, SP.)  
Livro de resumos da XII Semana Integrada do Instituto de  
Física de São Carlos/ Organizado por Adonai Hilario [et al.]. São  
Carlos: IFSC, 2022.  
446 p.  
Texto em português.  
1. Física. I. Hilario, Adonai, org. II. Título

ISBN: 978-65-993449-5-4                      CDD: 530

## PG132

# 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; AMBROSIO, André Luis Berteli

andre.jucovski@usp.br

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 (2): (i) a quantificação seletiva do componente dispersivo ( $\lambda$ -dependente) do fator de espalhamento atômico ou (ii) a substituição parcial do solvente aquoso ordenado por íons mais elétron-densos (metálicos ou halogênicos). 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). Agora têm-se como objetivo o refino de métricas para melhorar o desempenho de predição (erro médio de  $\sim 60^\circ$ ), além do desenvolvimento de uma avaliação quantitativa da fidelidade dos MDE.

**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.