Data analysis and interpretation in molecular medicine: from databases to artificial intelligence

Module 3: PROTEOMICS

July 14-15, 2021

Organizers

Enza Maria Valente (coordinator, PhD course in Translational Medicine)
Riccardo Bellazzi (former coordinator, PhD course in Bioengineering and Bioinformatics)

Objectives of the course

On day 1, participants will first have a background overview on "*protein science*" and learn about a wide variety of online databases for protein sequence analysis, compare them and learn how to use the most relevant functions. In particular, they will use the popular Uniprot database, learning how to "cherry pick" the information we want from thousands of results. They will, also, be introduced to commonly used online software for secondary structure analysis, protein properties analysis and post-translational modifications (PSIPRED, XTALPRED, Clustal, ExPasy ProtParam, NetNGlyc, TMHMM). Later, they will hear an overview of practical strategies to obtain a protein's structure (X-ray crystallography and cryo-electron microscopy), and then will be introduced to several methodological approaches to generate and analyze protein models starting from the amino acid sequence. In particular, they will discuss the main software to predict and visualize the 3D structure of a protein (SwissProt, Modeller, UCSF Chimer, PyMol). Finally, they will learn the basics of molecular dynamic simulations to predict the impact of amino acid changes on the structure and function of a protein.
On day 2, in the morning participants will focus on the use of strategies and tools for drug discovery, including a first section on molecular docking and structure-based approaches in drug discovery, and a second section on *in-silico* approaches for the identification of pharmacological targets. This will be illustrated also by means of practical examples. In the afternoon, the applicants will continue their training on drug discovery by focusing on the applications of statistical learning and deep artificial neural networks. A hands-on practice will be provided at the end of this section.

Instructors

Marco Lolicato – *Dept. of Molecular Medicine, University of Pavia*
Tommaso Mazza, Tommaso Biagini – *Bioinformatics Laboratory, CSS-Mendel Institute, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo and Rome*
Chiara Bianca Maria Platania – *Department of Biomedical and Biotechnological Sciences (BIOMETEC), University of Catania*
Toni Giorgino – *Biophysics Institute, Consiglio Nazionale delle Ricerche, Milano*

Venue

mixed (in presence max 30 attendants + online) – please specify your preference upon enrolling.

Course Program

Day 1

Morning session

9am – 9.30am Enza Maria Valente – Riccardo Bellazzi
Course introduction

9.30am-10.30am Marco Lolicato
Proteins and their properties
- *primary and secondary structure*
- *protein folding (tertiary and quaternary assembly)*
- *membrane proteins*

11.00am-12.30pm Marco Lolicato
Protein bioinformatic tools: databases and prediction software
- *From DNA to protein and protein properties (Expasy tools)*
- *Protein databases and boolean search tools (NCBI, Uniprot)*
- *Protein multiple alignments*
- *Secondary structure prediction (PSIPRED, etc.)*
- *Other useful tools*

Afternoon session

2.00pm-2.30pm Marco Lolicato
Introduction to structural biology
- *X-ray crystallography*
- *cryoelectron microscopy*

2.30pm-3.30pm Tommaso Biagini
From primary sequence to 3D model: structural bioinformatics
- *The Protein Data Bank: view on data and visualization*
- *Protein modeling*
- *Elements of Molecular Dynamics Simulation*

4pm-5.15pm Tommaso Mazza
Protein interaction
- *Data sources of Protein-to-Protein interaction*
- *Interaction networks and visualization*
- *Measures of topological importance*

Day 2

Morning session

9.30am-10.30am	Chiara B.M. Platania *Molecular docking and structure-based approaches in drug discovery: theory, tools and case studies*
11.00am-12.30pm	Chiara B.M. Platania *In-silico approaches for identification of pharmacological targets: case studies and practice* *MicroRNAs as pharmacological targets through their combinatorial effects on protein expression*

Afternoon session

2.00pm-3.00pm	Toni Giorgino *Statistical learning and deep artificial neural networks: introduction*
3.30pm-4.30pm	Toni Giorgino *Neural network architectures in drug discovery: selected examples*
4.30pm-5.00pm	Enza Maria Valente, Riccardo Bellazzi *Closing remarks*

Websites and Tools

https://academic.oup.com/gigascience/article/9/10/giaa115/5934256
https://ai-pharma.dka.global/
http://bioinf.cs.ucl.ac.uk/psipred/
https://blast.ncbi.nlm.nih.gov/Blast.cgi
http://www.cbs.dtu.dk/services/NetNGlyc/
http://www.cbs.dtu.dk/services/TMHMM/
https://www.cgl.ucsf.edu/chimera/
https://colab.research.google.com
http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=site/index
https://drugdesigndata.org/about/datasets
https://www.ebi.ac.uk/Tools/msa/clustalo/
https://www.expasy.org
https://genemania.org/
https://www.kaggle.com/c/MerckActivity
https://www.mirnet.ca/
http://moleculenet.ai/datasets-1
https://www.ncbi.nlm.nih.gov/gds
https://pyntacle.css-mendel.it/
https://www.pymol.org/2/
https://www.rcsb.org/
https://reactome.org/
https://salilab.org/modeller/
http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
http://www.sciences.univ-nantes.fr/elnemo/
https://string-db.org/
https://swissmodel.expasy.org/
https://thebiogrid.org/
https://tox21.gov/
https://www.uniprot.org
https://youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi
https://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl
https://zhanglab.dcmb.med.umich.edu/I-TASSER/

Software to be installed: cytoscape v. 3.7.0// excel (within cytoscape, install the following apps: Genemania and STRING)