MSc in Bioinformatics for Health Sciences

SBI. Structural Bioinformatics

Syllabus Information

Academic Course: 2018/19

Academic Center: 804 - Official Postgraduate Programme in Biomedicine

Study: 8045 – Bioinformatics for Health Sciences - MSc

Subject: 30169 – SBI. Structural Bioinformatics

Credits: 5.0

Course: 1st

Teaching languages: English

Teachers: Baldomero Oliva

Teaching Period: 2nd term

Presentation

The main goal of the course is to learn the basic concepts on the structure of macromolecules, more specifically of proteins. Also the principles of crystallography are given in order to gain a deeper understanding on the 3-dimensional data. The main features relating sequence and protein fold will be used on the prediction of secondary and tertiary structure of proteins and on its evaluation. Finally, the relationship between sequence/structure and function of proteins will be analyzed.

Associated skills

Instrumental:

1. Competence to unveil the relationship between the three-dimensional structure of bio-molecules and their biological activity
2. Proficiency on the main programs that handle macromolecular data of sequence and three-dimensional coordinates

Interpersonal:

1. Capacity to be integrated on a working multi-cultural team and split in milestones the analysis of biomolecules

Systemic:

1. Leading capacity on a working team and motivation to finish on time the work
2. Creativity and motivation to increase the quality.

Learning outcomes

Contents 1: Principles of Protein Structure

CONCEPTS	PROCEDURES	ATTITUDES
Concepts of protein structure: primary, secondary and tertiary	Visualization of protein structures in the computer	Organization of data
	Using DSSP to calculate secondary structures	
Classification of folds: all alpha, all beta, alpha/beta and alpha+beta	Identification of energetic terms involved in fold and folding pathway	Visual abstraction

Contents 2: Sequence comparison

CONCEPTS	PROCEDURES	ATTITUDES
Criticism to modify computer-based alignments	Use of the package HMMER and the PFAM database	Criticism to modify computer-based alignments
Differences between local and global alignments		
Methods of alignment: Needelman-Wunsch and Hidden Markov Models	Use of BLAST and PSI-BLAST	
Definition of sequence domains by PFAM	Use of ClustalW	
Multiple alignment of sequences		

Contents 3: Structure comparison

CONCEPTS	PROCEDURES	ATTITUDES
Structure superposition: pair-wise and multiple structures	Pair-wise and multiple structural alignment with STAMP	Visual abstraction on three-dimensional space
Calculation of backbone RMSD (root mean square deviation)		
Structure is more conserved than sequence		

Contents 4: Comparative Modeling

CONCEPTS	PROCEDURES	ATTITUDES
Distance restraints based on evolution	Use of MODELLER	Intuition to superpose structures and identify sequence conservation
Application of energies to optimize protein structure		
Protein flexibility on loops		

Contents 5: Threading and fold recognition

CONCEPTS	PROCEDURES	ATTITUDES
The Boltzman Law and relationship between energy and probability		Criticism to accept the solutions from fold recognition programs
Pseudo-energies and statistical potentials	Use of PROSA, THREADER, LOOPP, FUGUE, PHYRE and ModLink+	
Correlated mutations, mutual information and distance correlations theory	Rossetta approach	

Contents 5: Refinement and evaluation

CONCEPTS	PROCEDURES	ATTITUDES
Detection of modeling errors combining PROSA and PSIPRED: miss-alignments, wrong target, secondary-structure optimization.	Use of secondary structure prediction programs: PSIPRED	Criticism to modify sequence alignments and to refine a model
Helix capping	Use of the structural classification of capping: Arch-caps	
Structural classification of loops	Use of the structural classification of loops: ArchDB	
Relationship between secondary structure conservation and model refinement	Loop prediction: ArchPred and ModLoop	

Prerequisites

1) Knowledge in Bioinformatics: alignment algorithms (Needleman-Wunsch and Smith-Watermann) and substitution matrices (PAM and BLOSUM); Hidden Markov Models and Neural Networks; Linux and windows operative systems; Perl programming.

2) Knowledge in Mathematics: Derivatives and Integrals in n-dimensions; Taylor polynomial; Linear algebra and trigonometry; main concepts of statistics and probabilities: combinatorial analysis, permutations, Bayes theorem, Gaussian and Poisson distributions; numerical analysis, optimization and fitting of functions.

3) Knowledge in Physics: Newton classical mechanics; Thermodynamics, Gibbs energy, entropy, enthalpy and basic statistical mechanics.

4) Knowledge in Chemistry: Chemical structure, conformation, configuration and stereoisomers; Reactivity, transition state, electrophilic and nucleophilic attacks; Electronic density, molecular orbitals, quantum numbers, Schrödinger function and wave function; Steady state and equilibrium.

5) Knowledge in Biochemistry: 20 aminoacids, peptide bond, hydrogen bond and hydrophobic/polar properties of residues; Michaelis-Menten equation and enzyme kinetics and reactivity.

6) Knowledge in Molecular Biology: exonic structure and splicing; DNA duplication, transcription and translation; transcription factors; site directed mutagenesis; signalling pathways.

Contents

Theory Lectures Agenda

PART I: INTRODUCTION

Lesson 1. Introduction to the course. Design of the course. Distribution practices and requirements. Description of the work of structural biology. Evaluation and degrees.

PART II: PRINCIPLES OF STRUCTURAL BIOLOGY

Lesson 2. Proteins: polypeptide chain, secondary structure, tertiary and quaternary. Sequence and coding information. Conserved patterns by blocs. Sequence homology. Entropy and solvation. Energy principles: concepts of force and work. Entropic effect and solvent environment of clatrates. Thermodynamic principles. General principles of globular protein folding: hydrophobic core and secondary structure elements. Definition of secondary structure. The phi-psi space. The alpha helix and beta sheet. Supersecondary structure and connections of secondary structures (loops). Packing of alpha-helices. All alpha domains. Domains α / β: the Rossmann fold and TIM barrel. α + β domains. All beta domains: super barrel and β-barrel meanders, the sandwich greek-key and jelly roll.

PART III: CONFORMATIONAL SPACE

Lesson 3. Conformational space and molecular dynamics Molecular force fields. Using the potential energy function to find stable structures of three-dimensional models of proteins through minimization processes. The phase space, the conformational space and the partition function. Entropy (S), enthalpy (H) and free energy (G). Flexible systems and use of molecular dynamics to explore flexibility. Understand and apply computational methods such as the Mean-Force Potential and the free energy perturbation theory. Finite difference algorithms (Verlet). Simulated annealing. Solvent effect in simulations of proteins. Boundary conditions and treatment of the electrostatic field potential.

PART IV: STRUCTURE DETERMINATION

Lesson 4. COMPARATIVE MODELING The classification of proteins and evolutionary relationships. Definition of homology. Methods to superpose tertiary structures. Characterization of active sites and functional domains. Hidden Markov Models. PFAM and SMART databases. Alignment of sequences. Selection of the template. Detection of problems in the alignment. Variable regions and conserved regions. Classification of loops. Model building of the scaffold of a protein. Optimization of loops and side chains.

Lesson 5. Fold prediction The theorem of inverse folding. Statistical potentials. Neural networks and prediction of secondary structure (Threader and PSIPRED). Prediction methods of folding and threading. Inference of function (PHYRE and Modlink+). Alignment of secondary structure (TOPITS). Ab initio and mini-threading (Rosetta).

Lesson 6. Structures in Systems Biology Partition of protein domains. Interactions between chains and between domains. Predicting physical interactions based on domains. Transitive and permanent complexes. Other predictions of relationships between genes and proteins. Communication systems and signalling networks (phosphorylation). Study of interaction networks: Interactome.

Teaching methods

Week	Time	Extra time	Type	Subject	Program, Database
1	2h	1h	Theory	Protein Structure	
1	2h	2h	Practice	Protein Structure	SCOP, DSSP, RASMOL
2	2h	1h	Theory	Homology Modelling	
2	2h	2h	Practice	Sequence Alignment	PSI-BLAST
2	2h	2h	Practice	Sequence Alignment	HMMER, PFAM
3	2h	2h	Theory	Threading	
3	2h	3h	Practice	Structure Superposition	STAMP
4	2h	4h	Practice	Model Building	MODELLER
4	2h	4h	Practice	Model evaluation	PROSA
4	2h	4h	Practice	Threading	Threader, 3DPSSM, Fugue, LooPP
5	2h	2h	Practice	Model refinement	MODELLER
5	2h	2h	Practice	Modelling examples	MODELLER, PROSA
6	2h	2h	Theory	Protein interactions	
6	2h	3h	Practice	Docking	PatchDock, FTDock, ZDOCK, HEX
7	2h	2h	Practice	Model protein-protein interactions	MODELLER
7	2h	2h	Practice	Integrated examples	MODELLER, ZDOCK
8	2h	2h	Practice	Integrated examples evaluation	MODELLER, DOPE, ZRANK, TM_Score
9-12		48h	Project	Programming	Phyton

Evaluation

The evaluation is performed in three parts: theoretical exam (3/10), practical exam (3/10) and a work project (usually programmed in python) (4/10).

The **work project** implies groups of 2 students to generate a program of structural analysis. The aim of the program is to prove the relationships of structure, sequence and function, unveil the conservation of the most important residues and rationalize its structural role on the protein activity. The program can be standalone or in a web service and it is evaluated in association with the python course. It is mandatory that students not enrolled in the “Introduction to python” course form the group with one of the enrolled students (i.e. we ensure that the programs of all groups are also evaluated according to the lecturer of python). The mark of the project will consider the following criteria (weigths in parenthesis):

- Quality of the program: assessed by the lecturer of python (3)
- Executability: 0/1 if the program works (1)
- Installation: Easy to install (1)
- Quality of the tutorial: Explanation of the installation and execution with some examples (2)
- Theory : Theoretical background and methods developed (4)
- Applicability: 0/1 either the program can be applied to other problems or not (i.e. a database would always have 0) (1)
- Quality of the analysis: Explanations on examples and furtjher analysis of the quality of the results produced with the program (3)

The **practical exam** is aimed to prove the ability of the student to solve problems on molecular modeling and to characterize the protein structure

The **theoretical exam** is aimed to validate the basic knowledge of the student on protein folds.

Bibliography and Information Resources

Textbooks.

BRAND, Carl; TOOZE, John. Introduction to Protein Structure. 2a. ed. Garland Publishing, 1999.

Cantores; Schimmel. Biophysical Chemistry. WH Freeman & Co., 1980.

DAUNE, M. Molecular Biophysics. Oxford: University Press, 1999.

DRENTH. Principles of Protein X-ray Crystallography. New York: Springer-Verlag Inc., 1998.

Leach, A. Molecular Modelling: Principles and Applications. 2a. ed. Harlow: Pearson Education, 2001.

Finkelstein A, Ptitsyn O. Protein physics: a course of lectures. 2nd ed Academic Press. 2016.

PHILIP E. BOURNE Structural Bioinformatics / edited by Philip E. Bourne, Helge Weissig Hoboken, N.J. : Wiley-Liss, C2003

TAYLOR, W. R. (Willie R.). Protein Geometry, Classification, symmetry and topology: a computational analysis of structure / William R. Taylor and András Aszódi Bristol: Institute of Physics Pub., Once. 2005

Further reading

Ulloa LANGEL ET AL. Introduction to Peptides and Proteins / Langel by Ulloa, Benjamin F. Cravatt, Astrid Graslund, N.G.H. von Heijne, Matjaz Zork (2010)

A. PANCHENKO, T. Protein-protein interactions PRZYTYCKA and networks: computing methods for identification, analysis and prediction. London: Springer, 2008

PETER ALAN & Tompa FERSHT Structure and function of intrinsically disordered proteins / by Peter Tompa, Fersht Alan Publications. 2010

K. NAJARIAN Systems biology and Bioinformatics. CRC Press, 2009

Jake Y. CHEN & STEFANO LONARDI Biological data mining. 2009

Sheldon J. PARK Protein engineering and design. CRC Press, 2009

Gentleman, ROBERT, R Programming for Bioinformatics. Boca Raton, Fla. : CRC Press, c2009

Valiente Feruglio, Gabriel. Combinatorial pattern matching algorithms in Computational Biology using Perl and R. Chapman & Hall / CRC, 2009

BURKOWSKI, F. J. Structural Bioinformatics: an algorithmic approach. London: Chapman & Hall / CRC, c2009

Creighton, Thomas E. The biophysical chemistry of nucleic acids & proteins. Helvetian Press, 2010

IRENA ROTERMAN Structure-function relation in proteins / no. Standard 978-81-7895-409-7 b14339729

Teaching resources

http://sbi.imim.es/courses.php