Universitat Pompeu Fabra Barcelona

Course plan

2015-2016 Academic Year

Qualification Master's Degree

Structural Bioinformatics

Descriptive details concerning the subject:

- **Name of the subject:** Structural Bioinformatics
- **Code :** 30169
- **Type of subject:** Optativa
- **Module:** SBI
- **Credits:** 5 ECTS
- **Total hours:** 125.0
- **Year:** 1st
- **Term:** 2nd
- **Coordination:** Baldo Oliva
- **Department:** Ciències Experimentals i de la Salut
- **Building:** Campus Mar

Teaching staff:

· **Group:** 1
· **Language:** English
· **Lecturer:** Baldo Oliva
· **Department:** Ciències Experimentals i de la Salut

Subject presentation:

· 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 terciary structure of proteins and on its evaluation. Finally, the relationship between sequence/structure and function of proteins will be analysed.

Prerequisites in order to follow the itinerary:

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 polinomia; 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 estereoisomers; Reactivity, transition state, electrophilic and nucleophilic attacks; Electronic densitty, molecular orbitals, quantum numbers, Shroedingen 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.

Competences to be attained in the subject

Instrumentals:
1. Competence to unveil the relationship between the three-dimensional structure of bio-molecules and their biological activity
2. Competence to unveil the relationship between molecular patterns of aminoacids and nucleotides and their functions
3. 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 analyse 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

Evaluation

General assessment criteria:
The evaluation is performed in three parts: theoretical exam, practical exam and presentation of a work project.

The work project implies groups of 2, 3 or 4 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 has to be presented in a 10’ talk.

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.

Contents

Contents 1: Principles of Protein Structure

<table>
<tr><th>Concepts</th><th>Procedures</th><th>Attitudes</th></tr>
<tr><td rowspan="2">Concepts of protein structure: primary, secondary and tertiary</td><td>Visualization of protein structures in the computer</td><td rowspan="2">Organization of data</td></tr>
<tr><td rowspan="2">Using DSSP to calculate secondary structures</td></tr>
<tr><td rowspan="2">Classification of folds: all alpha, all beta, alpha/beta and alpha+beta</td><td rowspan="2">Visual abstraction</td></tr>
<tr><td>Identification of energetic terms involved in fold and folding pathway.</td></tr>
</table>

Contents 2: Sequence Comparison

<table>
<tr><th>Concepts</th><th>Procedures</th><th>Attitudes</th></tr>
<tr><td>Definition to score alignments between residues</td><td rowspan="2">Use of the package HMMER and the PFAM database</td><td rowspan="5">Criticism to modify computer-based alignments</td></tr>
<tr><td>Differences between local and global alignments</td></tr>
<tr><td>Methods of alignment: Needelman-Wunsch and Hidden Markov Models</td><td>Use of BLAST and PSI-BLAST</td></tr>
<tr><td>Definition of sequence domains by PFAM</td><td rowspan="2">Use of ClustalW</td></tr>
<tr><td>Multiple alignment of sequences</td></tr>
</table>

Contents 3: Structure Comparison

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

Contents 4: Comparative Modeling

Concepts	Procedures	Attitudes
Distance restraints based on evolution		
Application of energies to optimize protein structure	Use of MODELLER	Intuition to superpose structure and identify sequence conservation
Protein flexibility on loops		

Contents 5: Threading and fold recognition

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

Contents 6: Refinement and evaluation

Concepts	Procedures	Attitudes
Detection of modeling errors with PROSA: miss-	Use of secondary structure prediction programs:	Criticism to modify sequence alignments and to refine a

alignments, wrong target, secondary-structure optimization.	PSIPRED	model
Helix capping.	Use of the structural classification of cappings: 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.	

THEORY LECTURES AGENDA

PART I: INTRODUCTION
Item 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
Item 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
Item 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

Item 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.

Item 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).

Item 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.

Work Project

Groups of up to 4 students do the work project. It consists on programming a server or standalone program on the analysis of the structure of biomolecules. The project is evaluated in a 10’ presentation of the program and show of applications. Here is a list of suggestions, but it’s open to other programs or approaches under guidance and consulting with tutors or professors of the course.

1. Normal Modes analysis based on structures of the same family, structures from NMR, or molecular dynamics simulations.
2. Prediction of interactions by using Mirror Trees.
3. Correlated mutations and distance correlations to predict Aa interactions
4. Calculation of dipolar moments of biomolecules and in local regions.

5. Calculation of hydropaticity moments of biomolecules and in local regions.
6. Structure superposition algorithms of biomolecules
7. Extracting libraries of interfaces and clustering them
8. Superposition of interface regions
9. Development of statistic potentials for predicting structure.
10. Comparative analysis of normal modes between interacting and non-interacting conformations of two biomolecules
11. Classification of surfaces of local structures
12. Association between Kd and specific domain-domain interactions in PPIs of multidomain proteins.
13. Association between regions with Post-Transcriptional Modifications (PTM) and interfaces.
14. Structural classification of local structures with PTMs
15. Structural classification of active site conformations
16. Comparison between the network of CA distances and the network of Aa-sequence mutual information
17. Extracting protein-loops implicated in PPIs and DDIs
18. Automated modeling of PPIs based on templates
19. Comparison between Kd and structural properties of interfaces.
20. Modeling RNA structures.
21. Modeling and docking of RNA-protein interactions.
22. Tutorial web on the use of Rossetta Doc, HADDOCK, PyDOCK, HEX, FTDOCK, ZDOCK with several examples and assessment of models.
23. Tutorial web about the assessment quality of protein structures and Decoys analysis: QMEAN, Verify-3D, Prosa2003, DOPE, ANOLEA, MetaMQAPII, FoldX, ModFOLDclustQ, Multicom.
24. Package program for the comparison and evaluation of models of protein-structures (useful for CASP competition): RMSD, uRMSD, MaxSub Score, GDT_TS, 3D-Jury, TM-Score, Q-score.
25. Tutorial web on the use of IMP to reconstruct mixed models: Application on examples of TF complexes, Nucleopore, Spliceosome.
26. Tutorial web on the use of threading an ab-initio approaches with several examples: PHYRE, FUGUE, LOOPP, THREADER, Rossetta, Tasser, Direct Information.