Universitat Pompeu Fabra Barcelona

Year : 2018/19

3362 - Bachelor's (Degree) Programme in Human Biology
20431 - Structural Biology

Syllabus Information

Academic Course:	2018/19
Academic Center:	336 - Faculty of Health and Life Sciences
Study:	3362 - Bachelor's (Degree) Programme in Human Biology
Subject:	20431 - Structural Biology
Credits:	6.0
Course:	4
Teaching languages:	Theory: Grup 1: English Practice: Grup 101: English Grup 102: English Grup 103: English Grup 104: English Seminar: Grup 101: English Grup 102: English
Teachers:	Baldomero Oliva Miguel, Nuria Boada Centeno
Teaching Period:	Second Quarter

Presentation

Subject Name (20431) Structural Biology Qualification / Study: Human Biology Course: 4th Term: 2nd Number of ECTS credits: 6 Hours dedicated by the student: 66 class hours Language or languages of instruction: English Teachers: Baldomero Oliva, Nuria Centeno-Boada and Jana Selent

Presentation of the course

With this course the student acquire practical and theoretical knowledge about the structure of proteins and other biomacromolecules as well as the methods used for their characterization. The course includes: - The introduction to the biophysics of molecular systems. - The structural principles of biopolymers: proteins and DNA. - The three-dimensional structure determination of biomolecules. - The structure / function relationships of proteins. - Molecular simulation of proteins and DNA.

Associated skills

Competences to be achieved

1) Recognition of the structural patterns of biomolecules and relationship with

their biological function. The student must demonstrate understanding of the physicochemical descriptors of structure: terms of potential energy, solubility, acidity, hydrophobicity 2) Relating three-dimensional structure of biomolecules with their biological

function Demonstrate understanding of:

* Stability of proteins and determining their function. Principles of globular protein folding and membrane.

* Relationship between sequence, structure, and function: global and local flexibility and similarity of the sequence, three-dimensional preservation of active centres, conservation of interactions with ligands and other proteins.

* Bases and applications of homology model. Students should be able to:

* Recognize the types of folding and the main domains of proteins.

* Identify the conserved residues in structure and describe its possible structural function.

* Analyze the structural interaction between proteins or between proteins and other compounds.

3) Manage the basic software that allows processing data structures and

sequences of biomolecules. The student must demonstrate understanding of:

* The most important biomedical research current databases.

* Sequential Patterns, positional weight matrices, probabilistic models of sequences and hidden Markov models.

* Theoretical basis of protein homology modelling and ab initio model building Students should be able to:

* Using RASMOL program to visualize and analyze protein structures.

* Identify elements of a file of Protein Data Bank (PDB), extract the protein sequence and structure divided in their chains.

* Obtaining the secondary structure of a protein with DSSP and comparing it with a prediction (PSIPRED).

* To search for similarity using BLAST, FASTA and PSI-BLAST using multiple databases and matrices of weights, and in particular search for motifs in DNA sequences: splice sites, promoters reasons.

* Get a multiple sequence alignment with CLUSTAL and T-COFFEE

* Access and manage information from repositories of genomic and proteomic data: Ensembl, Genome Browser, NCBI, SWISSPROT, SCOP, CATH, PDB, PFAM and SMART.

4) To know the methods of experimental analysis of structures and biological

systems

The student must demonstrate understanding of:

* X-ray cristallography. Applications: analysis of a protein crystal. Principles of nuclear magnetic resonance (NMR). Two-dimensional NMR (applications to the structure determination of biomolecules).

Contents

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 4a. Protein crystallography and X-ray diffraction
Notions of wave mechanics. The principle of superposition of waves. Fourier series and Fourier transform. X-ray diffraction in crystals. Expression of the intensity of reflection. Bragg's law. Electron density maps. B-Factors and structural vibration. The phase problem. Multiple isomorphous replacement (MIR) and multiple anomalous diffraction (MAD). Structural refinement and molecular replacement.

Item 4b. NMR
The phenomenon of spin relaxation. NMR: the field of pulsed radiofrequency (RF) and the free-induction decay (FID). Nuclear Overhauser Effect (NOE). Two-dimensional NMR spectroscopy. Determination of three-dimensional structure of proteins and nucleic acids using distance constraints.
Item 5. 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 6. 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 7. 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.

PART V: STRUCTURE / FUNCTION

Presentation of 15 works chosen from the following topics (some topics may be subdivided). Requirements for the presentation are: 1) show the relevant interactions described in the bibliography using the programs learned during practices and theory; 2) show the details of specific hydrogen bonds, salt bridges and other kind of interactions relevant for the function of the protein (or RNA, DNA, etc.) and for their interactions with other macromolecules; 3) show the relevance of sequence similarities along evolution, indicating the specific functional regions and changes or conservation of Aas relevant to preserve or modify the function of the protein; 4) use and compare sequence alignments and structural alignments, analysing the different folds, superfamilies and families; 5) only if necessary, and with the agreement of the teachers, the group can also analyze models build by homology or threading to show, improve or implement a particular project. See some examples in the web of the structural bioinformatics group (sbi.imim.es); 6)

participation in the class is mandatory, comments of the presentation will be collected at the end of the class from all students present (additional 0,1 pts on the practical exam)
Structure of DNA and proteins related

The double helix of DNA(1). The major and minor grooves of the double helix. Forms B, and Z. The triple helix of DNA. Recognition of DNA bases.

Recognition of DNA: Transcription factors (2). Structure of the "lambda Cro" and its interaction with DNA: prediction by modeling of the interaction. Helix-turn-helix motifs, zinc fingers and leucine-Zippers. The transcription of DNA and transcription factors: family zinc-finger motifs. Homeodominis. Leucine-Zippers.

Packaging of DNA. Heterochromatin and Nucleosomes (3). Histones H1, H2, H3 and H4. Telomers and DNA stabilization in cell division (4).

Implication of ATP and NAD in the production of energy: PARP and poliadenineribose (5) , Nudix family (6).

Machinery of replication DNA polymerases (1), helicases (2), topoisomerase (3), etc..

DNA replication. Structure of DNA polymerase I and Klenow fragment. Structure of restriction enzymes and proteins stabilizing the single strand of DNA (SSB). Topoisomerase type I and type II and the release of torsional stress of DNA. Reverse transcriptase. The recombination of DNA: complex RuvA-RuvB-Holliday junction.

Translation machinery: RNA polymerase (1), Aa-t-RNA synthetase (2), the ribosome (3),

Structure of RNA polymerase from E. coli. Structure of aminoacyl-tRNA synthetases: the catalytic domain and the domain that identifies the anti-codon. The ribosome: 30S and 50S subunits of the prokaryotic ribosome. Docking of proteins in the ribosome structure of the 16S rRNA sequence. The combination of X-ray diffraction and high resolution electron microscopy.

Enzymes.

Processes of oxidoreduction using nucleotides NAD, FAD and FMN. The NAD-binding domain: the Rossmann fold and the hypothesis of Rossmann. Stereo-specífic proton transfer from NADH. The nucleotide binding motif. Fusion genes. FMN-binding motif. Cytochrome b5. Walker motifs A and B. Kinases: CDK, MAPK, and Ser-Tyr-kinase kinase. TIM Barrel and its origin.

Examples of catalytic enzymes: serinprotease (1), carboxypeptidase (2)
The catalysis by enzymes. The transition state, the active enzyme-substrate docking. Specific substrate-binding subsites. The serine-proteases: the catalytic triad of amino acids and oxyanion hole. Metalloproteasessuch as carboxypeptidases and metzincines.

Acetylases and deacetylases. Sirtuin histone deacetylases and aging mechanisms. Mechanisms of epigenetics.

Ig proteins of the immune system (1), MHC (2)

Somatic recombination involved in the variety of antibodies. Conserved and variable domains and complementarity determining regions. The antigen-antibody complex. Structure of the histocompatibility complex (MHC): Human Lymphocyte Antigen (HLA). Polymorphisms of the heavy chain of MHC class I.

Membrane proteins (rhodopsins, Porin) and theoretical and experimental methods

Difficulties in the crystallization of membrane proteins. Transmembrane helices and graphics hidropaticity using the sequence. Bacteriorhodopsin. Photosynthetic reaction center from Rhodopseudomonas viridis. Reaction of photosynthesis, energy transfer and tunneling. Structure of beta-transmembrane proteins: Porin.
Families of receptors and G proteins. G proteins and receptors for growth factors. The epidermal growth factor (EGF) and its receptor. The oncogene v-erb B. The insulin receptors, EGF and PDGF. Signal transduction: tyrosine-kinase. The signal amplification. Structure of adenilate-cyclase. Structure of a truncated form of p21 ras (cH-ras p21). The GTP binding motif: mutations involved in the oncogenic activation.

Nucleopore macrocomplex structure.

Theoretical and experimental methods for docking protein-protein interactions (1) and protein-DNA (2)

Comparison of complementary surfaces. Automated programs to fit protein-protein and protein-DNA. Combination of methods of fitting rigid molecules and molecular simulation methods.

RNA Structure and Spliceosome.

Types of links, packing and hydrogen bonding. Secondary structure. Prediction of RNA secondary structure. Structure of t-RNA: The T-loop and U-turns. Location anticodon. Formation of hybrids. Structure of ribozymes. Structure of RNase P. U1 components. snRNA. snRNP U1A. Protein C. A2F auxiliary factors. Structure of U2 and U5.

Capsids of viruses and infection mechanisms.

Icosahedral symmetry of spherical viruses. Using more than one polypeptide chain in a complex virus. The capsid of picorna viruses. Animal and plant viruses. Mechanism of infection by T2 bacterio-phage.

Cell structure scaffold.

Myosin, actin, vinculin and catenin in the mechanism of motility. Cadherin junctions. Microtubules, kinesin, dynein and katanin. Centromers and structure of the mitotic spindle.

Chaperons

Helpers of protein folding. Hsp 60, Hsp70, Hsp 90, Hsp100, GroEL and Groes. Role of the ATP-binding domain.

Ubiquitination system. Proteasome

Ubiquitination system and elimination of unfolded proteins. Structure of the proteasome. Structure of ubiquitin and the ring finger. Complex formation.

Comments to be collected after the presentation:

NIA:

TITLE OF THE PRESENTATION:

General comments:

Were sequence alignments correct?:

Were structural alignments correct?:

Was easy to follow the presentation?

Add your own question:

PRACTICAL PROGRAMME AND SEMINARS

The theory program is complemented by theoretical practice/seminars (two-hour sessions). The order of the presentation follows the order by which they are given in the calendar:

Seminar 1. Introduction to linux and graphics programs.

> Linux for beginners: command key, disc spaces. Browsing directories. Text editors.
> Visualization and analysis of biomolecules: the PDB file. Rasmol and VMD visualization of structures in PDB. Study of protein complexes. Splitting by chains. Exploration of interactions between proteins and molecules: Calculation of distances. Recognition of the interface and analysis of hydrogen bonding network. Use of DSSP.

Seminar 2. Characterization of motifs, secondary and tertiary structure.

> Characterization of structural motifs and supersecondary structures (beta-hairpin, alpha-hairpin, beta-alpha-beta motifs). Study of types of foldis: all alpha, all beta, alpha / beta and alpha + beta. Detailed study of the classification of domain structures and databases SCOP and CATH.

Seminar 3 Extracting information from protein structures.

> Principles of the structure determination by X-ray diffraction of proteins and biomolecules. The PDB: definitions and fomats.

Reading a PDB file.

Seminar 4 energy optimization and molecular dynamics simulation.

Principles of energy optimization. Function of energy, force field, minimitació methods (Newton-Rapson, Conjugate Gradient, steepest descent). Optimization of the energy of a protein.

Practice 1. Optimization of energy and molecular dynamics simulation.

Study the system (protein) and construction of the topology optimization of the structure. Molecular dynamics simulation model designed for homology. Analysis and evaluation of results of energy optimization. Analysis of the trajectory: Fluctuations, RMSD, core modes and hydrogen bonding.

Seminar 5. Protein function

How to show the association between sequence-structure and function.

Practice 2. Sequence alignment and remote homology search

Part 1: multiple alignments using ClustalW. Position Specific Substitution Matrix (PSSM). Search method using Psi-Blast.
Part 2: Hidden Markov models (HMMER). PFAM databases.

Practice 3. Superposition and structural characterization of folds

Overlapping proteins directed and RMSD calculation. Characterization of the folding of several types of problem structures using structural alignment programs (STAMP) and graphics programs. Superposition of structures.

Practice 4. Comparative Modeling.

Part 1: Searching for homologs to the protein problem and ClustalW multiple alignment.
Part 2: Alignment of structural homologues. Hidden Markov multiple alignment.
Part 3: Modelling with Modeller.
Part 4: Comparison between model and structural characterization of homologous structures with which this form was generated and optimized model. Using ProCheck.
Part 5: Optimization and validation of a stereochemical model.

Practice 5. Prediction of folding (threading).

Part 1: Analysing structures with statistical potentials of Prosa2003. Study of structural models generated by homology
Part 2: Predictive study of the structure of a protein using reverse threading (comparison with iTASSER, THREADER, PHYRE)
Part 3: Prediction of secondary structure prediction methods via the Internet and Machine Learning (PSI-PRED). Improving a homology model.

Seminar 6. Protein Docking.

Reviewing on the potential energy surfaces and functions of punctuation (scoring). Recognition and prediction of interaction sites (binding site).

Practice 6. Protein Docking.

Protein docking calculations with PyDock. Selection of binary complex structures. Macrocomplexes.

Evaluation

1. The evaluation of academic performance will be based on the following scale (about 10 total points):
a. Multiple-choice questions: 1.5 points
b. Essay questions: 2.5 points
c. Examination of practices: 3 points
d. Work presentation: 3 points
2. Any type of copy during the exams implies the FAILURE of the course.

3. To pass the course requires a 40% mark for each part. The final mark is obtained by averaging all parts with the corresponding weights. PASS requires a minimum of 5 out of 10.

4. Criteria to recover in case of failure

a. Exam retaking is only allowed to those who have done the practices and work-presentation. Those who cancelled their enrolment in the course cannot have it. Exam retaking will be in July.

b. Only essay and practice exams have retaking. Activities such as "work-presentation" maintain the score of the course.

c. Scores cannot be postponed to next year without being enrolled.

d. The work presentation has no retaking. The retaking tests for theory and practice are separate: students do not need to retake the parts that were already passed. For the retaken of the theoretical part, the mutiple-choice test is excluded and the grade is obtained with 40% of the theory essay, 30% of the work-presentation and 30% of the practical exam.

5. Extra points: extra points are added by practical assignments and assistance to the presentations, up to 1,5 points over 10 increase to the note of practical exam.

Bibliography and information resources

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, Alexei V. Protein physics: a course of lectures / Alexei V. Finkelstein, Oleg B. Ptitsyn AmsterdamBoston: Academic Press, cop. 2002.
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

5.2. 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

5.3. Teaching resources

http://sbi.upf.edu/courses.php