COURSE DESCRIPTION CARD - SYLLABUS

Course name

RNA structural bioinformatics [S2Bioinf2>RNA]

Course

Field of study
Bioinformatics

Year/Semester
1/1

Area of study (specialization)
–

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
15

Laboratory classes
30

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

4,00

Coordinators

prof. dr hab. inż. Marta Szachniuk
marta.szachniuk@put.poznan.pl

Lecturers

Prerequisites

The person starting the 2nd-cycle studies should have attained the learning objectives defined for the 1st-cycle studies in the resolution of the PUT Academic Senate - the learning objectives are available at the website of the Faculty of Computing and Telecommunications. The student starting this module should have the basic knowledge about structural biology, theory of algorithms, programming, data bases and combinatorial algorithms. Moreover, he / she should be able to solve the fundamental problems in the area of programming and analysis of biological data. In addition, with respect to the social skills, the student should show such attitudes as honesty, responsibility, perseverance, curiosity, creativity, manners, and respect for other people.

Course objective

1. The transfer of knowledge concerning (i) algorithms and optimization methods used to solve fundamental problems of modern biology, biochemistry, and bioinformatics of RNA structures, and (ii) currently best tools for acquisition, storage, processing and analysis of biological data concerning nucleic acids. 2. Introducing students to the architecture of the most popular tools used in RNA structural bioinformatics, including both early and contemporary ones, as well as the latest machine learning- based models. The emphasis is placed on pros and cons of these approaches, their limitations (often unnoticed by users of such systems), technologies they employ, quality and reliability of obtained results, and the methods for their validation. 3. Developing students' skills in mathematical modeling and solving computational problems of RNA structural bochemistry and biology using both, simple and advanced bioinformatic algorithms and models, as well as their ability to prepare data for deep learning models, train these models, and test new computational methods.

Course-related learning outcomes

Knowledge:

1. The student has an extended knowledge of the mathematical, statistical, optimization and artificial intelligence models used to represent structures of nucleic acid molecules on different levels of their architecture, their modeling, and analysis of processes involving these molecules.
2 The student knows methods, techniques and tools used in the process of solving complex tasks in the analysis of RNA structures, mainly of engineering nature.
3. The student knows an architecture and the functional details of the most popular structural data repositories and algorithms for structure processing (structure prediction, structure determination, structure evaluation and comparison, clustering, etc).
4. The student is familiar with issues in modeling and analysis of biological systems and structures based on theoretical foundations.
5. The student has a detailed knowledge of the computational modeling of structure prediction processes, structure comparison, evaluation and modeling.
6. The student has a detailed, based on solid theoretical foundation, knowledge of research planning, optimization and efficient algorithms used in RNA structural bioinformatics; knows the most important (from the historical point of view) and most popular algorithms to compute and predict structures, search for common substructures, align structures.
7. The student knows the trends in development of computational methods for crucial problems of structural biology and biochemistry of RNA, and is aware of the approaches proposed for solving novel problems in structural bioinformatics.

Skills:

1. The student uses and interprets the information acquired from the literature concerning the general problems of RNA structural bioinformatics and from specialist journals (scientific publications in Bioinformatics, BMC Bioinformatics, Nucleic Acids Research, Nature Methods, etc; internet services and portals from the bioinformatics area); interprets and evaluates their contents.
2. The student is able to use advanced techniques and computational tools for solving biological problems and can estimate their usefulness.
3. The student is able to apply knowledge from the area of biochemistry and related sciences to solve problems of RNA structural bioinformatics (for example, modeling, annotating, evaluating of structures).
4. Under the supervision of a teacher, the student can plan and solve research tasks (e.g. modeling of biomolecule structure with the use of bioinformatic tools, evaluating of RNA model quality and accuracy, clustering of structures based on their similarity) using the known computational methods.
5. The student applies mathematical methods as well as specialized IT techniques and tools to represent, process, and analyse structural data describing an architecture of nucleic acids on various levels of detail.
6. The student is able to design and develop computer software to solve selected structural bioinformatics problems using appropriate methods, techniques and tools.
7. The student prepares the presentation of obtained results in Polish and English (reports, multimedia presentations) and discusses them in the group.
8. The student can estimate the usefulness of new achievements of structural bioinformatics and biochemistry, especially in the area of processing and analysis of RNA structural data.

Social competences:

1. The student understands the necessity of learning (during the lectures students are introduced with

the latest achievements of structural bioinformatics; students have the possibility to present an interesting subject concerning the latest solutions in the field).
2. The student can co-operate and work in a group (laboratory gives a chance to work in groups).
3. The student understands the necessity of reading the scientific and popular journals in order to gain and extend bioinformatics knowledge.
4. The student has a creative attitude in professional and social activities.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Formative assessment
a) verification of assumed learning objectives related to lectures:
• answers to the questions regarding the material from previous lectures
b) verification of assumed learning objectives related to tutorials / laboratory classes:
• evaluation of the ability to perform the laboratory exercises
• evaluation of the report of the exercises (prepared partially during the laboratory and partially at home)
• evaluation of the exercises' execution during the laboratories.
Total assessment
a) verification of assumed learning objectives related to lectures:
• evaluation of the knowledge during the written test. The test is composed of 5 questions; the student can obtain 3 points for an answer to each question; 20% of the test is prepared as multiple choice questions; the remaining part has the form of an open questions; all questions concern the problems introduced during the lectures and practised during laboratory exercises. In order to get positive mark, one should obtain at least 8 points.
• discussion on the test results
b) verification of assumed learning objectives related to tutorials / laboratory classes:
• the verification of the skills and knowledge regarding the issues mentioned during the laboratories during the final written test
• the final evaluation is based on the weighted average, which includes the marks obtained for the projects, reports and activity during the laboratories.
Additional activities taken into account:
• discussion on the additional aspects of the topics,
• preparation of short presentation about a structural bioinformatics problem that was not presented by the lecturer,
• effectiveness knowledge gained during problem solving,
• remarks helping to improve the teaching materials or the teaching process.

Programme content

The lecture program covers the following topics. The first lecture provides a general introduction to RN structure bioinformatics, presenting key issues in this modern field along with characteristics of proposed solutions in terms of requirements, limitations, and computational complexity. The second lecture focuses on RNA secondary structure, its classification, representations, and modeling methods. Students learn about base pairing rules, common structural motifs, and methods for recording and visualizing RNA structures in various file formats. Classic and modern algorithms for secondary structure prediction are discussed, including those based on dynamic programming, thermodynamics, and deep neural networks, emphasizing the advantages, limitations, and accuracy of computational approaches. The third lecture explains experimental methods in structural biology such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy, as well as computational RNA 3D structure modeling techniques including template-based and deep learning approaches. The relevance of data quality, method limitations, and the role of computational prediction in RNA structural research are highlighted. The fourth lecture concentrates on visualization, representation, and annotation of 3D RNA structures, covering molecular visualization methods - from physical models to advanced computer graphics - as well as algebraic, geometric, and trigonometric structural representations. Key structural data formats and the basics of structural information encoding and interpretation are presented. The second part covers RNA 2D structure annotation based on atomic coordinates, base pair identification, pseudoknot classification, and topology visualization using contemporary tools. The next lecture addresses validation, comparison, and evaluation of RNA structural models, with special attention to topological issues such as knots and entanglements in RNA molecules. Students explore quality assessment

methods, stereochemical and topological validation tools, and the biological and computational aspects of RNA entanglements, including their detection in predicted and experimental structures using state-of-the-art computational methods. The latter part covers algorithms and quantitative metrics for comparing structures at various representation levels and practical use of such comparisons for prediction accuracy assessment and structural motif identification. The sixth lecture discusses methods for evaluating the accuracy of 3D RNA structure predictions, including geometric measures and optimal structure alignment algorithms. Advanced comparison methods based on contact analysis and topology are considered, incorporating torsion angles and deformation indices. The following lecture introduces Monte Carlo simulations and their applications in structural bioinformatics, explaining the origins of the method, differences between deterministic and stochastic simulations, and illustrating its use for simulating biological structures like RNA and protein macromolecular complexes through random sampling for model optimization. The lecture details the Metropolis algorithm and simulation procedures exemplified by macromolecular system modeling software. The subsequent lecture covers fundamentals of molecular mechanics and molecular dynamics, emphasizing equations of motion, potential functions, and atomic forces. Various force fields describing molecular interactions and their application in biomolecular modeling are detailed. The lecture includes an introduction to statistical mechanics and core simulation techniques like energy minimization, equilibration, and sampling, highlighting practical considerations such as boundary conditions and the importance of simulation parameters for reliable results. The final lecture is dedicated to student presentations, where volunteers showcase selected current problems and challenges in structural bioinformatics of RNA.
Laboratory exercises are strictly related to the lectures. The lecture is a theoretical introduction, while the laboratory allows for hands-on exercises related to the introduced topics and for implementation of the own solutions for simple bioinformatic problems.
Laboratory exercises are conducted as two-hour sessions in a computer laboratory and have a practical character, preparing students for independent use of available tools and programming libraries. The program covers a wide range of topics closely related to the lecture content. Students explore, among others, the canonical secondary structure of RNA, data formats, and definitions of structural motifs. They develop format converters with heuristic functions for resolving the order of pseudoknots and tools for motif identification. Students learn to predict RNA 2D structures based on sequence and alignment, compare structures, analyze non-canonical base pairs, and detect evolutionarily conserved RNA 3D modules. Exercises also include annotation of secondary structures using atomic coordinates, classical methods for comparing 3D structures such as RMSD and the angular MCQ measure, as well as molecular dynamics simulations - from the principles of running simulations, through analyzing results, to implementing their own trajectory analysis projects. The curriculum also plans sessions on structural modeling using the recent artificial intelligence models, including tasks related to predicting RNA 3D structures, predicting non-canonical pairings in 2D structures, and general quality assessment. Students learn to prepare datasets addressing issues like data redundancy and data leakage, and build automated data analysis and filtering pipelines based on solutions like RNA3DB. Moreover, exercises cover comparisons of AI methods (both shallow and deep learning) with classical methods to assess their effectiveness and the influence of different test datasets on computational outcomes. Students also undertake tasks using deep neural networks to detect conserved patterns at RNA-protein binding sites, aiming to deepen understanding of the function and structure of these biomolecules.

Course topics

The course topics cover issues related to obtaining data on RNA structures through computational and laboratory experiments, processing structural data at various levels of detail, analyzing structural features, data visualization, topology classification, structure annotation, motif searching, reference-based and reference-free structure assessment, detecting abnormalities in structural models, and modeling structures with or without experimental data. The course includes work with classical methods as well as deep learning models.

Teaching methods

1. Lectures: multimedia presentation, blackboard presentation illustrated with examples, multimedia show.
2. Laboratory classes: solving exercises, practical exercises, running computational experiments, discussion, group work, multimedia show, workshops, games, case study.

Bibliography

Basic:
1. J. Gu, P.E. Bourne, "Structural Bioinformatics"
2. P. Baldi, S. Brunak, "Bioinformatics: The Machine Learning Approach"
3. T. Schwede, M. Peitsch, "Computational structural biology. Methods and applications."

Additional:
Recent scientific publications from the area of structural bioinformatics.

Breakdown of average student's workload

	Hours	ECTS
Total workload	100	4,00
Classes requiring direct contact with the teacher	45	2,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	55	2,00