COMP598: Advanced Computational Biology Methods and Research

Fall 2016
School of Computer Science
McGill University

An introduction to computational structural and system biology. Theory and practice of Protein and RNA structure prediction, RNA-RNA, Protein-RNA, protein-protein interaction prediction, biological network analysis.

GENERAL INFORMATION

Syllabus:
This class extends the material covered in COMP462/561 (i.e. Computational Biology Methods and Research). We will introduce fundamental concepts and techniques in *computational structural biology and computational system biology*.

Computational structural biology aims to model and predict the structure of bio-molecules (e.g., proteins, DNA, RNA), and computational system biology provides methods to model and study biological systems. These fields of study has a broad impact on the understanding of gene regulation processes and biological system mechanisms. It also open new research directions in synthetic biology such as the design of new bio-molecular devices and the re-engineering of biological functions.

This class will cover the theory and practice of computational techniques in these research areas. Each session will feature a description of the methods and algorithms, and will be completed by a pratical session in which we will describe how to use state-of-the-art software.

Instructor:

- Prof. Jérôme Waldispühl (="http://www.cs.mcgill.ca/~jeromew/") (jerome.waldispuhl@mcgill.ca (mailto:jerome.waldispuhl@mcgill.ca)).

Teaching Assistant:

- Alexander Butyaev (alexander.butyaev@mail.mcgill.ca (mailto:alexander.butyaev@mail.mcgill.ca)).

Lectures:

- Tuesday & Thursday, 16:05am to 17:25pm in Strathcona Anatomy & Dentistry 1/12.

Office hours:

Prof. Jérôme Waldispühl	Tuesday 14:30-16:00	ENGTR 3106
Alexander Butyaev	Thursday 11:30-13:00	ENGTR 3110

Course Webpage: http://www.cs.mcgill.ca/~jeromew/comp598.html (http://www.cs.mcgill.ca/~jeromew/comp598.html)

Course Material:
All the material needed for this class will be available on the public course web page. There is no required textbook. Although, we recommend the following textbooks to deepen the material presented in class:

- [CB2000] Peter Clote and Rolf Backofen, *Computational Molecular Biology: An Introduction*, Wiley, 2000.
- [DEKM1998] Richard Durbin, Sean R. Eddy, Anders Krogh, and Graeme Mitchison, *Biological Sequence Analysis*, Cambridge University Press, 1998.
- [GR2014] Jan Gorodkin and Walter Russo, *RNA Sequence, Structure, and Function: Computational and Bioinformatics Methods*, Humana Press, 2014.

Lecture slides will be made available in PDF form on the course web page. Lectures will be recorded and available on MyCourse.

Evaluation
Your final grade will be calculated as follows:

- 30% for 2 assignments (15% each)
- 25% for the final exam
- 30% for the project report
- 15% for the paper presentation

The final exam is closed book and electronic devices are not allowed.

ANNOUNCES

Release Date	Due Date	Announce
Dec 9	Dec 12	An incomplete draft of a practice final exam (docs/practice_final_COMP598_F2016.pdf) is now available. The structure of the final exam will be the following: (1) 15-20 short questions (1-2 sentences, 40 points total); (2) 2 problems (30 point each). You will be allowed to use a 2 pages of notes. However, books and electronic devices will not be allowed.
Dec 9	Dec 12	The final exam is scheduled on Dec 12 at 9am in Building SH688 Room 295.
Dec 1		GROMACS 5 is now installed on SOCS servers. Please, read this documents with updated commands (docs/gromacs5-COMP598.pdf) if you are using it.

Release Date	Due Date	Announce
Nov 13	Dec 8 at 11:59pm	The second assignment (docs/HW2_COMP598_F2016.pdf) is now available.
Nov 6	Nov 8	You can now select your project online (https://goo.gl/forms/LSjduEy3GtL4NxyS2).
Nov 2	Nov 8	The descriptions of research projects (docs/comp598/projects/) are now available.
Oct 27	Oct 31	Assignment 1 (docs/HW1_COMP598_F2016.pdf) has been updated (One question went to bonus and threshold used in Q3 has been updated).
Oct 19	Oct 20	The class of Oct 20 has been cancelled.
Oct 16	Nov 15	Paper review assignment (docs/COMP598_review_F2016.txt) is now available.
Oct 6		The office hours of Alexander Butyaev are cancelled today.
Oct 1	Oct 14	A selection of paper to review is now available (data/COMP598). Please enter your preferences by October 14 in the online form (https://goo.gl/forms/QOcwyW1WXr8UAmb02).
Sep 29	Dec 12 at 9:00am	The final exam has been scheduled on December 12 at 9am. The location has not been confirmed yet
Sep 27	Oct 31 at 11:59pm	The first assignment (docs/HW1_COMP598_F2016.pdf) is now available.
Sep 13	Sep 14 at 1:30pm	Bioinformatics lab project presentation in Trottier 3120
Sep 6		Welcome to COMP598!

SCHEDULE

	Date	Topic	Material
Lecture 1	Sep 6	Syllabus. Introduction to RNA structure and function. Timeline of RNA bioinformatics.	[Slides] (docs/comp598_F2016_lecture_1.pdf)
Lecture 2	Sep 8	RNA minimum free energy secondary structure prediction. *Application:* The Vienna RNA package (RNAfold).	[Slides] (docs/comp598_F2016_lecture_2.pdf) Chapter 2 of [GR2014]

	Date	**Topic**	**Material**
Lecture 3	Sep 13	Boltzmann ensemble of RNA secondary structures.	[Slides] (docs/comp598_F2016_lecture_3.pdf) Chapter 4 of [GR2014]
Lecture 4	Sep 15	Rational sampling of RNA secondary structures. *Application:* The Vienna RNA package (RNAsubopt).	[Slides] (docs/comp598_F2016_lecture_4.pdf) Chapter 4 of [GR2014]
Lecture 5	Sep 20	Beyond thermodynamic models *Application:* The Rfam database.	[Slides] (docs/comp598_F2016_lecture_5.pdf) Chapter 3 of [GR2014]
Lecture 6	Sep 22	Alignment of structured RNAs *Application:* RNAalifold.	[Slides] (docs/comp598_F2016_lecture_6.pdf) Chapter 5, 8, and 9 of [GR2014]. Chapter 6, 9, and 10 of [DEKM1998].
Lecture 7	Sep 27	Simultaneous folding and alignment of structured RNAs *Application:* LocARNA	[Slides] (docs/comp598_F2016_lecture_7.pdf) Chapter 13 of [GR2014]
Lecture 8	Sep 29	Pseudo-knots and RNA-RNA interaction predictions. *Application:* The Vienna RNA package (RNAup).	[Slides] (docs/comp598_F2016_lecture_8.pdf) Chapter 19 of [GR2014]
Lecture 9	Oct 4	RNA 3D Modeling. RNA 3D modules and 3D motifs databases. *Application:* MC-Sym and the RNA 3D Hub	[Slides] (docs/comp598_F2016_lecture_9.pdf) Chapter 13 of [GR2014]
Lecture 10	Oct 6	Evolution of RNAs. RNA sequence/structure maps. RNA neutral network. *Application:* Jupyter notebook for simulation of the evolution of RNA populations (https://github.com/cgoliver/RNA-Popgen-Notebook).	[Slides] (docs/comp598_F2016_lecture_10.pdf) Chapter 16 of [GR2014]
Lecture 11	Oct 11	RNA inverse folding *Application:* The Vienna RNA package (RNAinverse) and IncaRNAtion.	[Slides] (docs/comp598_F2016_lecture_11.pdf) Chapter 16 of [GR2014]
Lecture 12	Oct 13	Introduction to Protein structure. Timeline of protein structure prediction. *Application:* The protein data bank (PDB)	[Slides] (docs/comp598_F2016_lecture_12.pdf)
Lecture 13	Oct 18	Protein secondary structure prediction using Neural Networks *Application:* PSIPRED	[Slides] (docs/comp598_F2016_lecture_13.pdf)
Lecture 14	Oct 25	Protein residue contact prediction *Application:* EVcoupling	[Slides] (docs/comp598_F2016_lecture_14.pdf)
Lecture 15	Oct 27	Protein Threading *Application:* RaptorX	[Slides] (docs/comp598_F2016_lecture_15.pdf)

	Date	Topic	Material
Lecture 16	Nov 1	Minimalist models: The HP lattice model *Application:* The CPSP software suite	[Slides] (docs/comp598_F2016_lecture_16.pdf)
Lecture 17	Nov 3	Molecular dynamics simulation *Application:* GROMACS	[Slides] (docs/comp598_F2016_lecture_17.pdf) [Tutorial] (docs/MD_tutorial.pdf) [Material] (http://amyloid.cs.mcgill.ca/MD.zip)
Lecture 18	Nov 8	Guest lecture by M. Smaoui ``Development Platform for Artificial Pancreas Algorithms''	
Lecture 19	Nov 10	Protein-Protein Interaction networks *Application:* Struct2net, IsoRank	[Slides] (docs/comp598_F2016_lecture_19.pdf)
Lecture 20	Nov 15	Paper Review: • Adrian Tousignant: Rivas & Eddy, "A Dynamic Programming Algorithm for RNA Structure Prediction Including Pseudoknots", Journal of Molecular Biology, 1999. • Jason Yao: Sato et al., "IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming", Bioinformatics, 2011. • Amir Kadivar: Sankoff, "Simultaneous Solution of the RNA Folding, Alignment, and Protosequence Problems", SIAM Journal of Applied Mathematics, 1985. • Jeremy Georges-Filteau: Boniecki et al., "SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction", Nucleic Acids Research, 2016. • James Bodzay: Adhikari et al., "De novo prediction of protein folding pathways and structure using the principle of sequential stabilization", Proceeding of the National Academy of Science, 2014.	
Lecture 21	Nov 17	Paper Review: • Matthew Cooke: Andronescu et al., "Computational approaches for RNA energy parameter estimation", RNA, 2010. • Bogdan Mazoure: Do et al.,"CONTRAfold: RNA secondary structure prediction without physics-based models", Bioinformatics 2006. • Pouriya Alikhani: Cruz & Westhof, "Sequence-based identification of 3D structural modules in RNA with RMDetect", Nature Methods, 2011. • Faizy Ahsan: Kazan et al., "RNAcontext: A New Method for Learning the Sequence and Structure Binding Preferences of RNA-Binding Proteins", PLoS Computational Biology, 2006.	

	Date	**Topic**	**Material**
Lecture 22	Nov 22	Paper Review: • Roman Sarrazin-Gendron: Reinharz et al., "Combining structure probing data on RNA mutants with evolutionary information reveals RNA-binding interfaces", Nucleic Acids Research, 2016. • Ron Davies: Mann et al., "CPSP-tools - Exact and complete algorithms for high-throughput 3D lattice protein studies", BMC Bioinformatics, 2008. • Paul Pereira: Nguyen et al., "Mapping RNA-RNA interactome and RNA structure in vivo by MARIO", Nature Communication, 2016. • Haji Mohammad Saleem: Kim et al., "Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights", Science, 2006. • Jaspal Singh: Chindelevitch et al., "Optimizing a global alignment of protein interaction networks", Bioinformatics, 2013.	
Lecture 22	Nov 24	Paper Review: • Julyan Keller-Baruch: Rousseau et al., "Classifying leukemia types with chromatin conformation data", Genome Biology, 2014. • Ramchalam K R: Cowen et al., "Predicting the Beta-Helix Fold from Protein Sequence Data", Journal of Computational Biology, 2002. • Navin Mordani: Remmert et al., "HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment", Nature Methods, 2012. • Thomas McAlear: Alushin et al., "High resolution microtubule structures reveal the structural transitions in αβ–tubulin upon GTP hydrolysis", Cell, 2015. • Hong Chuan Guo : Acevedo et al. "Mutational and fitness landscapes of an RNA virus revealed through population sequencing", Nature, 2014.	
Lecture 24	Nov 29	3D Genomics *Application:* 3DGB	
Lecture 25	Dec 1	Crowdsourcing molecular biology research *Application:* Phylo	

Note: This schedule is subject to modification.

RULES & POLICIES

Background & Pre-requisites

Good understanding of basic algorithms (equivalent to COMP251), and core molecular biology concepts (i.e. DNA, RNA, Proteins structure and function). A basic progamming in Python.
Undergraduate students: You can register to this class with the permission of the instructor. If you have not taken it yet, it is possible (and even recommended) to register at the same time to COMP462/561. *We will synchronize the material taught in COMP598 with the one of COMP462/561.*

Policy on discussion Board

Please follow common sense rules and etiquette for discussion board postings: be polite, avoid texting shorthand ("ur" instead of "you are", ...), choose a suitable subject line for your posting and use multiple postings for multiple subjects, keep your postings brief, etc.

Policy on collaborations

We greatly encourage you to discuss the assignment problems with each other. However, these discussions should not go so far that you are sharing code or giving away the answer. A rule of thumb is that your discussions should considered public in the sense that anything you share with a friend should be sharable with any student in the class. We ask you to *indicate on your assignments the names of the persons with who you collaborated or discussed* your assignments (including TA's and instructors).

Policy on re-grading

If you wish us to re-grade a question on an exam (or assignment), we will do so. However, to avoid grade ratcheting, we reserve us the right to re-grade other questions on your exam as well.

Policy on final grades

I will use the same rules and formula for calculating the final grade for everyone. We understand that your performances may be influenced by many factors, possibly out of your control. However, that is the only way we can be fair. The only exceptions will be medical exceptions. In that case, I will require a medical note, which has to be also reported to McGill, and to be informed as early as possible. Failure to comply to these rules, may results in the impossibility to invoke a medical exception.

Policy on Assignments

Due date/time, location/mode for returning your solutions, and accepted formats will be announced in class and indicated on the course web page.
Failure to return your assignment in time will results in penalties or even absence of grading. Late submission of 24h or less will receive a penalty of 20%. In all other cases, your assignment will be refused and not graded.
Importantly, solutions that do not follow the requested format will receive a penalty. By default, we only accept PDF or TEXT files. Images (if any) must be embedded in a PDF. Do not compress your files. All files must open on LINUX SOCS workstations.
The quality of the presentation of your solution is very important. Unreadable material, cryptic notations, or bad organization of the material will results in penalties, and potentialy even an absence of grading. If you scan your hand-written solutions, it is your responsability to ensure that you submit a high-quality image (i.e. excellent luminosity, contrast, focus and resolution).
The clarity of your explanations will also be an integral part of your final grade.

Policy on programming code

Questions in assignments may require you to write a Python program. We will provide, as much as possible, input and output data to test your programs. However, it will be your duty to ensure that your Java files compile on LINUX SOCS workstations.
We will not grade programs that do not compile on these machines.
Submission of class files (instead of Java source files) will be considered as an absence of submission. Do not compress your files.

Use of French in assignments and exams

In accordance with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

McGill policies

McGill University values academic integrity. Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offenses under the Code of Student Conduct and Disciplinary Procedures. See this link for more information.

CONTACT

If you have any additional question, you can contact the instructor:

Jérôme Waldispühl
⌂3630 University Street, Room 3106, Montreal QC H3A 0C6
✉ (E-mail) jerome.waldispuhl@mcgill.ca (mailto:jerome.waldispuhl@mcgill.ca)
☎ (Phone) +1 514 398 5018

Based on a template from BlackTie.co (http://blacktie.co)