Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Juan Cortés
Email	juan.cortes@laas.fr
Institution	LAAS-CNRS
Website	https://www.laas.fr/public/en
Group	Robotics and Interactions

Project

Web development & bioinformatic tools

Project Title:

A web server to generate conformational ensembles of highly-flexible proteins

Keywords:

Web server, Intrinsically Disordered Proteins (IDPs), Sampling, Conformational ensemble models

Summary:

Contrarily to what was thought in past decades, not all proteins fold into a relatively stable functional structure. Many proteins remain highly flexible in solution, possibly forming local transient structural elements. These proteins are usually called Intrinsically Disordered Proteins (IDPs). They play crucial roles in multiple biological processes and are directly involved in several pathologies, including cancer and neurodegeneration. The high flexibility of IDPs has notably hampered their study. Experimental biophysics technics such as Nuclear Magnetic Resonance (NMR) and Small-Angle X-ray Scattering (SAXS) provide information on conformational trends [1]. However, the quantitative interpretation of these experimental data requires the use of computational approaches that account for their ensemble averaging properties. These computational approaches are based on the construction of large conformational ensembles. We have recently developed a new method to model conformational ensembles of IDPs [2], which provides a more accurate representation than existing approaches. This method will be of great inserts for the scientific community working on the understanding of IDPs. The goal of this project is to provided easy access to this method through a web server, as we did a few years ago for another molecular modeling application (http://moma.laas.fr) [3]. By the end of the project, aiming to disseminate our work, we plan to submit an article describing the new web server for publication in a high-impact scientific journal. The student will work in a team involving other students (PhD and master level), researchers and software engineers working on related topics. He/she will take part in the design phase and the full-stack web development (both front-end and back-end). We aim to use the most recent languages and technologies at both levels (in particular, the Django web framework). Particular importance will be given to the ergonomy of the proposed solution.

References:

[1] T.N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estaña, J. Cortés, N. Sibille, P. Bernadó (2017) Small-angle scattering studies of intrinsically disordered proteins and their complexes. Current Opinion in Structural Biology, 42:15-23. [2] A. Estaña, N. Sibille, E. Delaforge, M. Vaisset, J. Cortés, P. Bernadó (2019) Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. Structure, 27(2):381-391 [3] D. Devaurs, L. Bouard, M. Vaisset, C. Zanon, I. Al-Bluwi, R. Iehl, T. Siméon, J. Cortés (2013) MoMA-LigPath: a web server to simulate protein-ligand unbinding. Nucleic Acids Research, 41(W1):W297-W302.

Expected skills::

Good programming skills are mandatory, mainly C++ and Python. Teamwork skills are also very important for the achievement of the project.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Jana Selent
Email	jana.selent@upf.edu
Institution	IMIM-UPF
Website	www.jana-selent.org
Group	GPCR Drug Discovery Group

Project

Web development & bioinformatic tools

Project Title:

Developing web-based analysis tools for the study of GPCRs

Keywords:

web-based tools, data-sharing, structural biology, G protein coupled receptors

Summary:

G protein–coupled receptors (GPCRs) are major targets for the pharmaceutical industry (more than 30% of all FDA-approved drugs act on a GPCR) and present an immense potential for future drug development. Although GPCRs have been extensively studied over the past decades, the underlying molecular and structural mechanisms responsible for many critical regulatory processes of this protein superfamily remain elusive. Understanding the dynamics of receptor functionality is currently a major challenge in molecular biophysics and a requirement for the rational design of drugs with improved therapeutic profile. For this reason, a promising approach to elucidate the molecular basis of GPCR functionality are molecular dynamics (MD) simulations, a potent computational technique capable of generating atomic-resolution simulations of the structural motions of a molecular system. Consequently, MD simulations are increasingly being applied to the study of GPCRs, as reflected by the rapid upsurge of publications concerning this topic. In view of the growing importance of MD simulations, the GPCRmd project was created with the purpose to build the GPCRmd database, a database of MD simulations of GPCRs capable to foster data from all-over the world (www.gpcrmd.org). This web-based platform provides visualization and analysis tools specifically designed for the evaluation of structural and dynamic data of GPCR family members. In this Master project, the student will be involved in the design and development of new interactive analysis tools for the study of GPCRs, which will be incorporated in the GPCRmd viewer webpage. Such analysis tools will be focused on the study of the complex signalling network of intra-protein interactions that ultimately determine the response of the GPCR to a given drug. This includes the automatized detection and classification of different types of relevant interactions, comparison of the interaction network of different receptors (phylogenetically related receptors, wild type vs. mutant, ...), comparison of the network at different stages of a given molecular process, etc. Moreover, the obtained analysis tools will be applied to a case-study with the final aim to better understand how the intra-protein interaction network of a receptor of interest is affected by different stimuli or alterations (binding of a given ligand, receptor mutations, ...). The student will learn about GPCR biology, protein dynamics and in silico drug design, as well as web development, biomedical data analysis and biological databases. The internship can be extended into a PhD thesis. We expect that the GPCRmd database will have high impact on GPCR research and the discovery for new drugs. This will be communicated in a relevant publication to which the Master student will contribute.

References:

https://www.biorxiv.org/content/biorxiv/early/2019/12/17/839597.full.pdf

Expected skills::

Experience in structural biology. Python, HTLM/CSS and web page design. Experience with molecular dynamics simulations and JavaScript is a plus. Good level of English (oral and written).

Possibility of funding::

To be discussed

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Davide Cirillo
Email	davide.cirillo@bsc.es
Institution	BSC - Barcelona Supercomputing Center
Website	https://www.bsc.es/
Group	Computational Biology

Project

Web development & bioinformatic tools

Project Title:

Evolution and crosstalk of biological ontologies

Keywords:

biological ontologies; graph theory; machine learning

Summary:

The research undertaken at the Barcelona Supercomputing Center (BSC) by the Computational Biology group, led by Prof. Alfonso Valencia, covers a wide range of Artificial Intelligence approaches for biomedicine, in particular in the area of biomedical computational graph theory and algorithms with special focus on biological ontologies. Biological ontologies, such as the Human Phenotype Ontology (HPO) (Köhler et al. 2019) and the Gene Ontology (GO) (The Gene Ontology Consortium 2019), are recognized as essential tools in the grand challenge of biomedical data integration and interpretation. In collaboration with the BSC Computer Science Department, we have developed a system for the efficient traversing and exhaustive path enumeration in interconnected biological ontologies (Cirillo et al. 2019) exerting large-scale parallelism and scalability in High-performance computing (HPC). This framework harnesses machine learning to infer a precise mapping between disease-related phenotypic features and distinct molecular processes allowing knowledge discovery. The proposed activity will be centered on the application and extension of this framework to a larger set of biological ontologies with the aim to study aspects such as (1) the dynamics of biological knowledge accumulation across time; (2) the integration and reconciliation of the multiple biological ontologies; (3) the implementation of machine learning approaches for biological knowledge representation and reasoning. The selected candidate will work in a highly sophisticated HPC environment, will have access to systems and computational infrastructures, and will establish collaborations with experts in different areas. What will you learn - Computational biology: biological knowledge representation; resources, formats and tools related to ontologies for use across the biomedical domain; applications and analytical approaches based on ontological information. - Computer Science: basics of High-performance computing; use of BSC supercomputing resources; BSC biology-oriented HPC implementations. - Scientific Dissemination: acquisition of science communication skills through lab meeting presentations and research article writing.

References:

Köhler et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019 Jan 8;47(D1):D1018-D1027. doi: 10.1093/nar/gky1105. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330-D338. doi: 10.1093/nar/gky1055. Cirillo et al. Graph analytics for phenome-genome associations inference. bioarXiv. 2019 Jun 26. doi: https://doi.org/10.1101/682229.

Expected skills::

- Good statistical and programming skills (Python, R/Bioconductor) - Strong interest in the analysis of biological systems - Basic knowledge of bioinformatics and molecular biology - Ability to access and evaluate scientific literature

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

This project will mainly focus on the application and extension of a previously developed tool, which is currently used in the laboratory. The student will be in close contact with collaborators at the BSC Computer Science Department within the groups “Best Practices for Performance and Programmability” led by Javier Teruel Garcia and Marta Garcia Gasulla, and “High Performance Artificial Intelligence” led by Ulises Cortés. The project will be supervised by Davide Cirillo and co-supervised by Alfonso Valencia.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Esteban Vegas i Ferran Reverter
Email	evegas@ub.edu; freverte@ub.edu
Institution	University of Barcelona
Website	
Group	Estadística i Bioinformàtica

Project

Web development & bioinformatic tools

Project Title:

Deep Learning based approaches for mining biomedical databases

Keywords:

Deep Learning, Data mining, Databases, Natural language processing

Summary:

This project aims the implementation and development of a tool based on Deep Learning models for the extraction and abstraction of biomedical knowledge using machine learning analysis of the contents in biomedical references databases (PubMed, MedGen, ...). Specific searches for terms related to a target disease will feed deep clustering algorithms to determine a set of disease-related descriptors. Then, recurrent-neural networks must be trained to assign automatically biomedical articles to disease-related descriptors.

References:

COHEN, Aaron M.; HERSH, William R. A survey of current work in biomedical text mining. Briefings in bioinformatics, 2005, vol. 6, no 1, p. 57-71. Gully A Burns, Xiangci Li, Nanyun Peng, Building deep learning models for evidence classification from the open access biomedical literature, Database, Volume 2019, 2019, baz034, https://doi.org/10.1093/database/baz034 Lan, K., Wang, D., Fong, S. et al. A Survey of Data Mining and Deep Learning in Bioinformatics. J Med Syst 42, 139 (2018). https://doi.org/10.1007/s10916-018-1003-9

Expected skills::

Python, Machine Learning, Databases, Data mining.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Josep F Abril
Email	jabril@ub.edu
Institution	Department of Genetics, Microbiology & Statistics
Website	https://compgen.bio.ub.edu/
Group	Computational Genomics Lab @ UB

Project

Web development & bioinformatic tools

Project Title:

Integration of omics data related to retinal dystrophies

Keywords:

Interaction networks, retinitis pigmentosa, interologs mapping, web interface, graph databases

Summary:

Inherited retinal dystrophies (IRDs) comprise a highly heterogeneous group of disorders caused by over 200 causative genes. The prevalence of IRDs is 1:3000 worldwide, which make these blinding disorders a health relevant target. The implementation of massive sequencing approaches has greatly facilitated genetic testing and, as a result, the number of IRD genes and mutations is constantly increasing. Nonetheless, a substantial number of cases remain to be accurately diagnosed, as the average yield in IRD genetic diagnosis is roughly 50%. Our lab has contributed to the implementation of the RPGeNet (https://compgen.bio.ub.edu/RPGeNet) network and the curation of RNA-seq data to project differential gene-expression over that interaction network. The main goal is to perform computational analyses, integrating further omics data, to explore and pinpoint novel candidate genes and pathways that can be associated to molecular components of the disease. We plan to extend the current interaction network based on human genes/proteins, into further model organisms, such as mouse and zebrafish, in a way that the network of interologs can facilitate the integration of expression data from the three organisms.

References:

RPGeNet2 publication: https://academic.oup.com/database/article/doi/10.1093/database/baz120/5618821

Expected skills::

Student should master Unix/bash, python/C/perl, R, and perhaps some SQL. We will introduce the student into graph databases and django to provide interactive access to the data.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

If applicant is interested in doing a PhD, there is the possibility to apply for a Generalitat FI or Ministerio FPU PhD grants on the next announcement.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	María José Rementería
Email	maria.rementeria@bsc.es
Institution	BSC - Barcelona Supercomputing Center
Website	https://www.bsc.es/
Group	Social Link Analytics

Project

Web development & bioinformatic tools

Project Title:

Dynamics of spreading false health news on social networks [RRSSalud]

Keywords:

Health, Fake News, Social Media

Summary:

Summary The RRSSalud is a research project that aims at investigating the typology and dynamics of dissemination on social media of fake news in the area of health. By combining quantitative methodologies (statistical analysis and social network analysis) and qualitative techniques (content analysis and focus groups), we explore the attitudes of people regarding the health information they consume. Specifically, we plan to focus on understanding the ability of Internet users to distinguish between false and true content, as well as, the tactics and strategies they use to detect the trustworthiness of news. In addition, topics, morphologies and rhetorical strategies of fake news will be explored. The investigation is coordinated by three teams of researchers from the University of Navarra and the Barcelona Supercomputing Center. As a result, we expect to develop tools, methods, and guidelines that can be employed by public health institutions, media organizations, and the general public to counteract the dissemination of fake news. Introduction In 2017, the Royal Spanish Academy incorporated a new word into the Dictionary of the Spanish Language: post-truth, which is defined as the "deliberate distortion of a reality, which manipulates beliefs and emotions in order to influence public opinion and social attitudes." The incorporation into the language of this neologism is a symbol of a serious current problem: the organized dissemination of false information, mainly through social media, in order to manipulate the public opinion. This phenomenon, popularly known as false news or "fake news" [1], has shown to have a significant impact in multiple areas and situations in recent years. In the political arena, for example, it has been found that it significantly influenced the results of the Brexit referendum and the 2016 presidential elections in the United States [2]. In business, the phenomenon of disseminating hoaxes and biased information has been identified as a way to deliberately discrediting brands and companies [3]. Information on the environment has also been a fertile area of lies and half-truths, especially with regard to the information on climate change [4]. However, the area of health is one of the areas where disinformation can cause profound damages [5], with anti-vax campaigns or health recommendations on epidemic periods to name a few. RRSSalud project focuses on studying fake news in health-care published on social media in Spain. The aim of the project is to explain the relationship between the vulnerability of Internet users to fake news and the dynamics of propagation and repercussion of these contents on social media. By combining experimental research with quantitative and qualitative methods we plan to explore the phenomenon from a holistic and comprehensive perspective. Specifically, the goals of the project are: i) identify the typology of fake news in the area of health and disseminated on social media in Spain; i) assess the vulnerability of Spain’s Internet users to fake news; iii) profile population groups in relation to their critical capacity to identify fake news; iv) understand the dynamics dissemination of fake news and propose actions to mitigate its impact; v) identify the subjective aspects that lead to giving credit to fake news and promote its subsequent dissemination by users.

References:

1 Quandt, T., Frischlich, L., Boberg, S., & Schatto - Eckrodt, T. (2019). Fake news. The International Encyclopedia of Journalism Studies, 1-6. 2 Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. Social Science Computer Review, 37 (1), 38-54; Rose, J. (2017). Brexit, Trump and Post-Truth Politics, Public Integrity, 19 (6), 555-558. 3 Berthon, P. R., & Pitt, L. F. (2018). Brands, truthiness and post-fact: managing brands in a post-rational world. Journal of Macromarketing, 38 (2), 218-227. 4 Kolmes, S.A. (2011). Climate change: a disinformation campaign. Environment: Science and Policy for Sustainable Development, 53 (4), 33-37. 5 Viviani, M., & Pasi, G. (2017). Credibility in social media: opinions, news, and health information — a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7 (5), e1209.

Expected skills::

Experience with software development and at least one of the following: Proficiency in Python programming language / Proficiency in web programming languages and frameworks (e.g., Javascript, HTML, CSS, Django/Flask) // Ideally, some experience with Python data science tools (e.g., Pandas, Numpy, Jupyter Notebooks, Scikit-learn, Matplotlib/Seaborn) to obtain, curate, clean, analyze, and visualization of information // Ability to work in an interdisciplinary social-tech environment and interact with relevant stakeholders to understand their needs and formulate solutions

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

RRSSalud is one of the five projects funded by BBVA Foundation as part of its program Scientific Research Teams in Economics and Digital Society. The work will be supervised by Nataly Buslón, Jorge Saldivar, and Maria José Rementeria.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Carles Hernandez-Ferrer and Leslie Matalonga
Email	carles.hernandez@cnag.crg.eu, leslie.matalonga@cnag.crg.eu
Institution	CNAG-CRG
Website	www.cnag.cat
Group	Data Analysis Team - Bioinformatics Unit

Project

Web development & bioinformatic tools

Project Title:

Implementation of an automated genomic (re)analysis support system for rare disease diagnostic.

Keywords:

rare diseases, genome-phenome analysis

Summary:

It is estimated that 350 million people worldwide suffer from one of the approximately 7000 existing rare diseases (RDs). As 80% of RDs are thought to have a genetic origin, particular emphasis has been placed on the rapidly expanding development of genomic technologies. However, the interpretation of the genome is still a real challenge for molecular geneticists, and innovative bioinformatics solutions combining genomic and clinical data are crucial for reaching accurate patient diagnosis. At the CNAG-CRG and in the context of several EU projects (RD-Connect, Solve-RD, EJP-RD) we developed a data sharing and analysis tool, the RD-Connect Genome- Phenome Analysis platform (GPAP: https://platform.rd-connect.eu/), to provide methods and standardised analyses of phenotypic and (gen)omic data in order to facilitate mutation detection within the context of rare diseases. As part of the Solve-RD project: “solving the unsolved rare diseases” (www.solve-rd.eu), rare disease patient data from 19.000 genomic datasets will be reanalysed using the RD-Connect GPAP. To this purpose, high throughput SNV-indel data (re)analysis solutions are being implemented in the system to automatically allow real-time queries to a high number of samples. A first prototype has been built enabling the filtering of thousands of genomic datasets by specific filters including variant pathogenicity and population databases. The master student joining this project will be working on the development of this innovative approach by improving the tool that queries the RD-Connect API to enable more complex queries (involve family members’ genotypes, specific regions of the genome, other individuals with phenotypic similarity, etc.). This work will be done in close collaboration with clinicians and researchers of four ERNs (European Reference Networks) involved in the Solve-RD project enabling continuous analysis feedback and molecular diagnosis confirmation.

Expected skills::

Python programming Data base architecture Genetics background Team working skills

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Marc Güell

Email	marc.guell@upf.edu
Institution	UPF
Website	https://www.upf.edu/en/web/synbio
Group	Translational Synthetic Biology

Project

Web development & bioinformatic tools

Project Title:

Computational approaches for efficient and safe engineering of human genomes

Keywords:

Gene editing, CRISPR, Synthetic Biology, Gene therapy

Summary:

Our laboratory is focused on applied synthetic biology for therapeutic purposes. We have two lines of research, one in technology development for gene therapy, and one in skin microbiome engineering. Advanced cell and gene therapies are gaining important impact in medicine. We currently have more than 2,500 on-going gene therapy trials on multiple diseases (cancer, genetic disease, infectious disease, etc...). However, multiple concerns have been raised on the safety of current technologies which prevent a wider deployment. Uncontrolled on-target, pro-cancer pathway activation, controversy on off-target, and lack of efficacy still represent a major concern. We are offering a Bioinformatics master thesis position in developing computational approaches to develop more precise technologies for gene editing. AI and genomics provided tools to significantly improve design (Chuai et al, Genome Biology 2018; Doench et al, Nat Biotech 2016; ...). Nevertheless, these approaches remain not predictable enough for therapeutic purposes. We are developing new algorithms which use clinically relevant data to improve prediction significance for therapeutic purpose.

Expected skills::

Basic statistics and programming

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Master project 2020-2021

Personal Information

Supervisor	Gianni De Fabritiis
Email	gianni.defabritiis@upf.edu
Institution	UPF
Website	https://www.compscience.org/
Group	Computational Science Laboratory - GRIB

Project

Web development & bioinformatic tools

Project Title:

Abstraction and reasoning challenge: Create an AI capable of solving reasoning tasks it has never seen before

Keywords:

reinforcement learning; machine learning; inductive programming; AI

Summary:

Can a computer learn complex, abstract tasks from just a few examples? Current machine learning techniques are data-hungry and brittle—they can only make sense of patterns they've seen before. Using current methods like reinforcement learning, an algorithm can gain new skills by exposure to large amounts of data, but cognitive abilities that could broadly generalize to many tasks remain elusive. This makes it very challenging to create systems that can handle the variability and unpredictability of the real world, such as domestic robots or self-driving cars. However, alternative approaches, like inductive programming, offer the potential for more human-like abstraction and reasoning. The abstraction and reasoning corpus (ARC) provides a benchmark to measure AI skill-acquisition on unknown tasks, with the constraint that only a handful of demonstrations are shown to learn a complex task (https://www.kaggle.com/c/abstraction-and-reasoning-challenge). This competition was initially created by the creator of the Keras neural networks library and it’s explained in this paper (https://arxiv.org/abs/1911.01547). The idea is to move beyond the competition timeframe to create an AI that can solve reasoning tasks it has never seen before and set up a path toward a PhD in AI. It is expected that novel work in terms of a paper should be produced during this period. For further details, contact Gianni De Fabritiis (gianni.defabritiis@upf.edu). The research period is paid. We are looking for exceptional candidates passionate about AI and with the willingness to go beyond in AI research. The lab is very well equipped.

References:

http://grib.imim.es/publications/index.php?CATEGORY1=14

Expected skills::

Be confident with maths and programming

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Master project 2020-2021

Personal Information

Supervisor	Gianni De Fabritiis
Email	i.escolar@acellera.com
Institution	Acellera Labs
Website	https://www.acellera.com/
Group	-

Project

Web development & bioinformatic tools

Project Title:

Machine learning in computational structural biology and drug discovery

Keywords:

machine learning; GPU computing; medicnal chemistry for drug discovery; PlayMolecule

Summary:

This project aims to develop machine learning methods applied to structural biology, drug discovery and computational chemistry. The aim is to go substantially beyond the state-of-the-art in the use of machine learning and GPU computing, exploring supervised, unsupervised and reinforcement learning approaches. We expect the candidate to participate in the development of new learning approaches and applications derived from deep learning applied to medicinal chemistry for drug discovery. By working in this project, the researcher will have access to state of the art computational resources. This project is expected to lead to discoveries that will be publishable in the highest impact scientific journals. Some examples of applications can be seen in PlayMolecule.org, a drug discovery platform used by thousands of scientists worldwide and pharmas and biotech companies. The platform is based on two main pillars, physical-based molecular simulations on GPUs and machine learning/AI, thus contributing to the company mission of accelerating the transition towards computerized drug discovery process. The platform was born in 2017 and serves as a repository of web applications for molecular modelling tools such as ProteinPrepare [Martínez-Rosell2017; doi:10.1021/acs.jcim.7b00190] and pioneering deep learning applications such as Kdeep [Jiménez2018; doi:10.1021/acs.jcim.7b00650] WHO WE ARE: Founded in 2006, Acellera was one of the first companies worldwide to leverage the

use of novel accelerator processor technology (GPU) for molecular simulations. Among our clients, we count 10 of the top 50 pharmaceutical companies. We were selected as one of the Top30 AI Drug Discovery companies in 2019. Our software includes PlayMolecule, ACEMD, HTMD, KDEEP, etc. and it's used by hundreds of users both in academia and the private sector. In particular, PlayMolecule.com is the first platform to democratize the use of molecular dynamics and machine learning applications for drug discovery.

Expected skills::

You have VERY good programming skills and a background in either chemistry, biology, computer science or similar Prior knowledge in neural information processing, deep learning frameworks (pyTorch, Tensorflow) is desirable Very good knowledge of Python and good coding practices This is a strongly computational position, so we encourage application of people that love algorithms, computing, programming and likes to apply it

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

What we offer: - Real in-company work experience. - Possibility to participate in the development of a real software product - After graduation, it is possible to stay in the company

---

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Gianni De Fabritiis
Email	g.defabritiis@acellera.com
Institution	Acellera Labs SL
Website	www.acellera.com
Group	-

Project

Web development & bioinformatic tools

Project Title:

Development of frontend capabilities and graphic user interface (GUI) for PlayMolecule.org

Keywords:

frontend; GUI, in silico drug discovery; HTML, Phyton

Summary:

This master thesis will focus on the development and release of a new graphic user interface (GUI) for PlayMolecule, a popular platform for biomolecular-related applications with hundreds of unique users every month available at https://www.playmolecule.org. PlayMolecule is a drug discovery platform used by thousands of scientists worldwide and pharmas and biotech companies. The platform is based on two main pillars, physical-based molecular simulations on GPUs and machine learning/AI, thus contributing to the company mission of accelerating the transition towards computerized drug discovery process. The platform was born in 2017 and serves as a repository of web applications for molecular modelling tools such as ProteinPrepare [Martínez-Rosell2017; doi:10.1021/acs.jcim.7b00190] and pioneering deep learning applications such as Kdeep [Jiménez2018; doi:10.1021/acs.jcim.7b00650]. The position is based within the informatic and innovation hub of Barcelona (Spain). Acellera values excellence, merits and scientific innovation above anything else. WHAT YOU WILL BE WORKING ON: 1. You will closely work with Acellera developers to develop and improve the GUI capabilities for PlayMolecule. 2. You will mostly work in front-end tasks such as: - Collaborate with medicinal chemists to improve the capabilities and usability of PlayMolecule web interface - Development of 3D novel graphical interface to better access molecular structural information and dynamic plots ready to interface with the PlayMolecule back-end. 3. You may additionally work in back-end tasks such as: - Development of functionalities necessary to make custom GUI molecular selections WHO WE ARE: Founded in 2006, Acellera was one of the first companies worldwide to leverage the use of novel accelerator processor technology (GPU) for molecular simulations. Among our clients, we count 10 of the top 50 pharmaceutical companies. We were selected as one of the Top30 AI Drug Discovery companies in 2019. Our software includes PlayMolecule, ACEMD, HTMD, KDEEP, etc. and it’s used by hundreds of users both in academia and the private sector. In particular, PlayMolecule.com is the first platform to democratize the use of molecular dynamics and machine learning applications for drug discovery.

Expected skills::

THIS PROJECT IS FOR YOU IF: You have good programming skills and a background in either chemistry or computer science You are proficient in: HTML, CSS/CSS3, Javascript, AngularJS, Python And maybe also have some knowledge of: Flask, SQL databases, Plotly.js, NGL.js You have very good communication skills in English

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

WHAT WE OFFER: - Real in-company work experience. - Possibility to participate in the development of a real software product - After graduation, possibility to stay in the company.

Master project 2020-2021

Personal Information

Supervisor	João Curado
Email	joao.curado@flomics.com
Institution	Flomics Biotech
Website	www.flomics.com
Group	Liquid biopsies group

Project

Web development & bioinformatic tools

Project Title:

Deconvolution of cell-free RNA transcriptome using RNA-seq

Keywords:

Plasma RNA; deconvolution model; sample heterogeneity; diagnosis;

Summary:

Years of literature demonstrate the existence of cell-free RNA, both messenger RNAs (mRNAs) and long noncoding RNAs (lncRNAs) originating from a wide variety of organs, from heart to the brain, and which change in response to external stimuli, namely diseases. Cell-free RNA molecules, circulating in human fluids such as plasma, saliva or urine, are potential windows into the health, phenotype or development stage of a variety of human organs, in a minimally invasive way. Despite this huge promise, the use of RNA sequencing (RNAseq) methods for global profiling of cell-free RNAs is in its infancy. In this project we propose to take advantage of the public available databases of RNA-seq such as Genotype-tissue expression (GTEx) consortium and tissue-specific gene expression databases, to determine the relative RNA contributions of each tissue in a sample using different methods (quadratic programming, least-squares regression, etc.). From a standard plasma RNA-seq experiment, the resulting tool will be used to calculate the relative contributions of the tissues and to monitor unexpected abnormalities that can be used as warning signs for complex disease detection.

References:

Koh, W. et al. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. Proc. Natl. Acad. Sci. 111, 7361–7366 (2014). Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457 (2015). Everaert, C. et al. Performance assessment of total RNA sequencing of human biofluids and extracellular vesicles. Sci. Rep. 9, 17574 (2019).

Expected skills::

Transcriptomics; statistics; programming; autonomy

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed