Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Carolina Estarellas
Email	cestarellas@ub.edu
Institution	Universitat de Barcelona
Website	http://www.ub.edu/cbcg/team/
Group	Computational Biology, Chemistry & gastronomy

Project

Structural bioinformatics

Project Title:

Exploring the molecular basis of the selective activation of specific AMPK isoforms by direct activators

Keywords:

protein dynamics - structure - activation mechanism - MD simulations - interaction networks

Summary:

Up to 30 % of all human proteins could be modified by protein kinases, which are known to regulate most of cellular pathways. The mammalian adenosine monophosphate-activated protein kinase (AMPK) is a Ser/Thr protein kinase that has an important role in cellular energy homeostasis and its activity is tightly regulated by the cellular ratios between AMP, ADP and ATP. AMPK is a 145kDa heterotrimeric complex which consists of a catalytic α subunit and two regulatory subunits, β and γ. Multiple genes in mammals encode all three subunits. The evolutionary adaptation of AMPK to different tissues is accomplished through the expression of distinct isoforms that can form up to 12 αβγ complexes, which exhibit notable differences in the sensitivity to allosteric activators. Due to AMPK central role in energy homeostasis, its activity is susceptible to being regulated by several mechanisms. These are: i) The binding of AMP to the CBS domains in the γ-regulatory subunit promotes the phosphorylation of Thr172 in the activation loop at the kinase domain of the α -subunit by upstream kinases, promoting an allosteric activation. ii) The indirect AMPK activators that act by increasing the cellular AMP concentration, such as metformin, phenformin, or oligomycin. iii) In 2006, Abbott Laboratories reported a novel mechanism of action that involves the first direct activation of AMPK by the thienopyridone drug A-769662. In contrast to adenine nucleotides, A-769662 does not bind to the CBS motifs in the γ-subunit but to a binding site located at the interface between α and β subunits which is called Allosteric Drug and Metabolite binding (ADaM) site. In the last years several direct activators of AMPK were reported. Parts of these activators are specific for certain isoforms of AMPK; some of them can activate both β1 and β2 isoforms while the others can only trigger the activation in one specific isoform. To shed light into the molecular determinants of the allosteric regulation of AMPK, we have already examined the structural and dynamical properties of β1- and β2-containing AMPK complexes formed with A-769662 and SC4 activators trying to dissect the mechanical response leading to active-like enzyme conformations through the analysis of interaction networks between structural domains. The results of these analyses show the mechanical sensitivity of α2β1 complexes in contrast to the large resilience of α2β2. Our results indicate that the binding of the activator to α2β1 promotes the pre-organization of the ATP-binding site, favoring the adoption of the activated form of the enzyme. Moreover, we hypothesize that the change of β1Asn111 to β2Asp111 could be the key factor in modulating the mechanical sensitivity of β1 and β2 containing AMPK complexes. However, still several questions remain to be answered at different levels. Firstly, we have performed the study considering different β isoform, but the same α isoform. So, a needed step is the understanding of the influence of α1 and α2 isoforms over the activation mechanism. Moreover, due to the complexity of the system the AMPK structures crystallized until now are not fully complete, and therefore it has not been simulated the full heterotrimeric complex. So, how the study of the full complex could affect the allosteric mechanism between the different subunits? How could this fact affect the different isoform composition? These questions are key factors that we need to tackle in order to complete the puzzle and shed some light in the drug design for specific isoform complex.

Expected skills::

1) Understanding the protein structure and noncovalent interactions, 2) Use of linux operating system, 3) Preferably: knowledge on visualization programs like pymol, vmd, 4) Preferably: knowledge on MD simulations and some packages to run MD simulations like amber, gromacs, etc

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Jordi Juárez
Email	jordi.juarez@ub.edu
Institution	University of Barcelona
Website	
Group	Computational Biology and Drug Design Group (CBDD)

Project

Structural bioinformatics

Project Title:

Molecular Basis of Cooperativity in Protein-Ligand-Protein complexes

Keywords:

PPI glues, cooperativity, molecular dynamics

Summary:

Multicomponent systems are extremely common in biochemistry: from allosteric modulation to organelle assembly, the underlying principle is the simultaneous interaction of more than two partners. While there are examples in which the affinity between partners is purely additive, it is often the case that these affinities are extremely affected by the presence of the remaining members of the multicomponent complex. This phenomenon is known as cooperativity and albeit been widespread and impacting many biological processes, its molecular determinants remain poorly understood. [1.2] A particularly relevant case of cooperative effects is the ability of

some small molecules to stabilize the interaction between macromolecules, acting as so-called molecular glues. The stabilization effect is often orders of magnitude higher than what one may infer from the affinity of the molecular glue for any of the macromolecular partners and this mechanism is often exploited by nature to modulate biological responses. In particular, stabilizers of protein-protein interactions (PPIs) are commonly used by different organisms to control their cellular functions and regulate signaling pathways. Additionally, molecular glues are also powerful chemical probes to help us better understand the function of macromolecular complexes, and even a few noteworthy examples (e.g: cyclosporin A, Paclitaxel or Rapamycin) have been developed into drugs. However, our lack of understanding of the cooperativity phenomenon has to date prevented the purposeful and reliable development of molecular glues. In fact, the discovery, optimization and development of most of these compounds have vastly relied in serendipity and/or trial and error assays. We propose to push a paradigm shift into the development of molecular glues, moving away from trial an error and into rational design. A critical step is to understand the molecular events that underpin PPI stabilization. Employing enhanced sampling techniques and data analysis tools[3,4], the student will seek to characterize the mechanism at play during ternary complex formation in selected examples from the literature and attempt to provide a general framework to describe cooperativity in biomolecular systems.

References:

[1] de Vink, P. J. et al. Cooperativity basis for small-molecule stabilization of protein-protein interactions. Chem Sci 10, 2869-2874, doi:10.1039/c8sc05242e (2019). [2] Andrei, S. A. et al. Stabilization of protein-protein interactions in drug discovery. Expert Opin Drug Discov 12, 925-940, doi:10.1080/17460441.2017.1346608 (2017). [3] Juárez-Jiménez, J. et al. Dynamic design: manipulation of millisecond timescale motions on the energy landscape of cyclophilin A. Chemical Science 11, 2670-2680, doi:10.1039/C9SC04696H (2020). [4] De Simone, A. et al. A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. Chem Sci 10, 542-547, doi:10.1039/c8sc03831g (2019)

Expected skills::

Familiarity with the AMBER software package will be an advantage

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

We are willing to consider other projects in the field of Structure Based Drug Design that match your interests and skills. Do not hesitate to get in contact.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Salomé Llabrés
Email	salome.llabres@ub.edu
Institution	Universitat de Barcelona
Website	http://www.ub.edu/cbcg/index.php

Group CBCG

Project

Structural bioinformatics

Project Title:

Molecular mechanisms underpinning active drug efflux in Gram-negative bacteria

Keywords:

Membrane Proteins, Molecular Dynamics Simulations, Antimicrobial Resistance

Summary:

Antimicrobial resistance (AMR) threatens to rise morbidity, mortality and treatment cost of microbial infections in the near future. The recent increase in multidrug-resistant strains compromises the current therapeutic arsenal against bacterial infections. The case of Gram-negative (GN) bacteria is especially alarming, as listed by the World's Health Organisation for R&D of new antibiotics [1]. Active drug efflux is one of the principal mechanisms driving both intrinsic and acquired multidrug resistance in GN bacteria. It relies on the expression of tripartite efflux pumps (TEPs), which are specialised protein complexes spanning both GN membranes able to expel drugs from the bacterial cytoplasm and periplasm [2]. Since the inhibition of TEPs promises to restore the efficacy of a wide range of existing drugs, there have been a lot of efforts to characterise the components and the complex assemblies of these bacterial machinery. Leveraging the available structural information obtained by X-ray crystallography and CryoEM tomography, we aim to obtain an atomistic description of the drug export processes of TEPs. Using a combination of homology modelling, unbiased molecular dynamics (MD) simulations and enhanced sampling techniques, we will identify the molecular mechanisms underpinning the substrate export from the periplasm and the cytosol via the inner membrane transporter to the outside of the bacterial cell.

References:

[1] WHO (2017) List of bacteria for which new antibiotics are urgently needed [2] Du D et al. (2018) Nat Rev Microbiol 16, 523-539

Expected skills::

Experience with physics- and structure-based methods (e.g. Molecular dynamics) and MD software packages (e.g. GROMACS) will be an advantage.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Comments:

We are willing to consider other projects in the field of biomolecular simulations that match your interests and skills.

Master project 2021-2022

Personal Information

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

Project

Structural bioinformatics

Project Title:

Unraveling signaling bias at G protein-coupled receptors (GPCRs)

Keywords:

G protein-coupled receptors, molecular dynamics, data analysis, drug design

Summary:

G protein-coupled receptors (GPCRs) are the most abundant class of receptors in the human organism. They are present in almost every type of cell, and govern almost every process in the human body (i.e. cognitive and inflammatory processes or control of the cardiovascular system). Owing to their ubiquity, they are targets of more than 30% of current drugs, and every day new GPCRs are revealed to be pharmacological targets for existing diseases. GPCRs can initiate signalling through binding with several intracellular partners, which in turn elicit distinct signalling cascades. It is established that in physiological conditions, a receptor binds to each of the partners in equal proportion. Interestingly some drugs act by altering the GPCR structure so that it preferentially binds to one specific partner - a phenomenon known as signaling bias. Such ligands, named biased ligands, offer great promise, as they enable to modify pathways associated with symptoms while not modifying other pathways - which could cause side effects. However, the molecular requirements for a molecule to act as a biased agonist within a receptor are still poorly understood. Molecular dynamics (MD) is a novel and sophisticated technique that enables to simulate protein behaviour in a physiological environment. The Master student will apply this approach to study time-resolved molecular mechanisms underlying signaling bias induced by small drug-like molecules. For this, the student will be trained on setting up simulated systems, running production simulations as well as the application of a wide range of analysis tools that allow capturing subtle structural events related to signaling bias. Structural insights will be exploited for the design of a novel class of GPCR modulators with a tailored signaling profile. We expect that the results of the analysis will be published in a high impact journal, and the expertise acquired by the student will make her/him a valuable asset for pharma companies in future. We are looking for a highly motivated and skilled student with exceptional academic records that allows pursuing a PhD afterwards.

References:

Rodríguez-Espigares & Torrens-Fontanals et al. GPCRmd uncovers the dynamics of the 3D-GPCRome, Nature Methods 2020, DOI: 10.1038/s41592-020-0884-y

Expected skills::

Experience in structural biology, programming in python/bash, molecular dynamics engines (GROMACS, NAMD, etc.), analysis tools/packages (VMD, Chimera, MDtraj…) and high 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 2021-2022

Personal Information

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

Project

Structural bioinformatics

Project Title:

Detection of novel druggable binding sites at G protein-coupled receptors

Keywords:

G protein-coupled receptors, molecular dynamics, data analysis, drug design

Summary:

G protein-coupled receptors (GPCRs) are the most abundant class of receptors in the human organism. They are present in almost every type of cell, and govern almost every process in the human body (i.e. cognitive and inflammatory processes or control of the cardiovascular system). Owing to their ubiquity, they are targets of more than 30% of current drugs, and every day new GPCRs are revealed to be pharmacological targets for existing diseases. Molecular dynamics (MD) is a sophisticated technique that enables to simulate protein behaviour in a physiological environment. In this project the Master student will develop a simulation-based pipeline that allows detecting druggable binding sites including cryptic pockets ("transient binding pockets"). This pipeline involves (i) the setup of simulation systems including small chemical fragments to probe the entire protein surface for druggable binding sites, (ii) running production runs, (iii) automated detection of binding sites and (iii) intuitive visualization. The pipeline will be implemented into our GPCRmd server and detected sites will be exploited for the discovery of new molecular GPCR modulators. We expect that the results of the analysis will be published in a high impact journal, and the expertise acquired by the student will make her/him a valuable asset for pharma companies in future. We are looking for a highly motivated and skilled student with exceptional academic records that allows pursuing a PhD afterwards.

References:

Rodríguez-Espigares & Torrens-Fontanals et al. GPCRmd uncovers the dynamics of the 3D-GPCRome, Nature Methods 2020, DOI: 10.1038/s41592-020-0884-y

Expected skills::

Experience in structural biology, programming in python/bash, molecular dynamics engines (GROMACS, NAMD, etc.), analysis tools/packages (VMD, Chimera, MDtraj…) and high 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 2021-2022

Personal Information

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

Project

Structural bioinformatics

Project Title:

Simulation meets experiment: refinement of cryo-EM GPCR structures using molecular dynamics

Keywords:

G protein-coupled receptors, molecular dynamics, data analysis, drug design

Summary:

G-protein coupled receptors (GPCRs) are the most abundant class of receptors in the human organism. They are present in almost every type of cell, and govern almost every process in the human body (i.e. cognitive and inflammatory processes or control of the cardiovascular system). Owing to their ubiquity, they are targets of more than 30% of current drugs, and every day new GPCRs are revealed to be pharmacological targets for existing diseases. Recent advances in cryo-electron microscopy (cryo-EM) and image classification provide insights into ensembles of low- to high-resolution that describe differently populated conformational states of proteins. However, methods for deriving accurate atomistic models from cryo-EM density maps lag behind this resolution revolution. The increasing amount of molecular detail requires the development of new methodologies and software to accurately and timely interpret experimental densities. Molecular dynamics (MD)-based refinement methods have grown into a valuable approach to tackle this challenge. In this project, the Master student will develop an MD-based pipeline that can be applied to GPCRs. This represents an important milestone for the scientific community as it can provide novel structural insights into this important drug targeting class. For this, the student will learn how to setup simulations and how to use correlation-driven MD for the refinement of atomistic models into cryo-electron microscopy maps. We expect that the results of the analysis will be published in a high impact journal, and the expertise acquired by the student will make her/him a valuable asset for pharma companies in future. We are looking for a highly motivated and skilled student with exceptional academic records that allows pursuing a PhD afterwards.

References:

Rodríguez-Espigares & Torrens-Fontanals et al. GPCRmd uncovers the dynamics of the 3D-GPCRome, Nature Methods 2020, DOI: 10.1038/s41592-020-0884-y Igaev et al. Automated cryo-EM structure refinement using correlation-driven molecular dynamics, Elife 2019, DOI: 10.7554/eLife.43542

Expected skills::

Experience in structural biology, programming in python/bash, molecular dynamics engines (GROMACS, NAMD, etc.), analysis tools/packages (VMD, Chimera, MDtraj…) and high 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 2021-2022

Personal Information

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

Project

Structural bioinformatics

Project Title:

Why is water essential for GPCR functionality?

Keywords:

G protein-coupled receptors, molecular dynamics, data analysis, drug design

Summary:

G protein-coupled receptors (GPCRs) are the most abundant class of receptors in the human organism. They are present in almost every type of cell, and govern almost every process in the human body (i.e. cognitive and inflammatory processes or control of the cardiovascular system). Owing to their ubiquity, they are targets of more than 30% of current drugs, and every day new GPCRs are revealed to be pharmacological targets for existing diseases. As for the whole life, water is critical for GPCR functionality at a nanoscopic scale. It forms important intermolecular networks that mediate the signaling response of the receptor. Our group owns an extraordinary dataset with unprecedented information about the implication of water molecules in GPCR structural dynamics and the binding of small drug-like molecules. We believe that stabilized or disrupted intermolecular water signatures drive the functional consequences of a drug-like molecule. In this project, the Master student will seek for the underlying molecular mechanism of the water-mediated functional responses. For this, she/he will (i) develop an analysis pipeline to extract relevant information from our unique dataset and (ii) setup control simulations to validate obtained conclusions. We expect that the results of the analysis will be published in a high impact journal, and the expertise acquired by the student will make her/him a valuable asset for pharma companies in future. We are looking for a highly motivated and skilled student with exceptional academic records that allows pursuing a PhD afterwards.

References:

Rodríguez-Espigares & Torrens-Fontanals et al. GPCRmd uncovers the dynamics of the 3D-GPCRome, Nature Methods 2020, DOI: 10.1038/s41592-020-0884-y

Expected skills::

Experience in structural biology, programming in python/bash, molecular dynamics engines (GROMACS, NAMD, etc.), analysis tools/packages (VMD, Chimera, MDtraj…) and high 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 2021-2022

Personal Information

Supervisor	Xavier Barril
Email	xbarril@ub.edu
Institution	Universitat de Barcelona
Website	http://www.ub.edu/bl/
Group	Computational Biology and Drug Discovery

Project

Structural bioinformatics

Project Title:

Enhanced Sampling Molecular Dynamics for Virtual Screening Applications

Keywords:

Computer-aided drug discovery, virtual screening, binding kinetics, binding free energy, molecular dynamics

Summary:

The goal of the Barril's lab is to discover first-in-class bioactive molecules, with special emphasis on non-standard mechanisms of action (e.g. allosterism, protein-protein glue). To do so, we apply state of the art structure-based drug discovery methods, many of which have been developed in-house. We introduced the use of Dynamic Undocking (DUck), a very efficient form of steered molecular dynamics.[1] The method can provide a fast approximation to protein-ligand dissociation rates (k-1),[2] reveal the structural stability of the complex,[3] or be used to predict binding modes.[4] Most importantly, it is sufficiently fast that it can be used as a post-docking tool in virtual screening (VS) applications, removing up to 80% of docking false positives.[1] To further improve the VS success rate, we are developing a complementary tool (LigBinder), that will be applied in cascade, after docking and DUck. LigBinder is a particular implementation of biased molecular dynamics to provide a fast estimate of the protein-ligand association rates (k1). In this project you will validate LigBinder, first for on-rate prediction, comparing with experimental data.[5] Then, in prospective VS applications. This project is synergistic with other projects in our lab and will benefit from substantial previous work and close collaboration with other group members.

References:

1. Ruiz-Carmona, S., Schmidtke, P., Luque, F. J., Baker, L., Matassova, N., Davis, B., Roughley, S., Murray, J., Hubbard, R., & Barril, X. (2017). Dynamic undocking and the quasi-bound state as tools for drug discovery. Nature Chemistry, 9(3), 201–206. https://doi.org/10.1038/nchem.2660 2. Schmidtke, P., Luque, F. J., Murray, J. B., & Barril, X. (2011). Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. Journal of the American Chemical Society, 133(46), 18903–18910. https://doi.org/10.1021/ja207494u 3. Majewski, M., Ruiz-Carmona, S., & Barril, X. (2019). An investigation of structural stability in protein-ligand complexes reveals the balance between order and disorder. Communications Chemistry, 2(1). https://doi.org/10.1038/s42004-019-0205-5 4. Majewski, M., & Barril, X. (2020). Structural Stability Predicts the Binding Mode of Protein-Ligand Complexes. Journal of Chemical Information and Modeling, 60(3), 1644–1651. https://doi.org/10.1021/acs.jcim.9b01062 5. Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H.-P., Dreyer, M. K., Frech, M., Lowinski, M., Vallee, F., Bianciotto, M., Rak, A., & Wade, R. C. (2018). Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. Journal of Chemical Theory and Computation, 14(7), 3859–3869. https://doi.org/10.1021/acs.jctc.8b00230

Expected skills::

molecular dynamics, structure-based drug discovery

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

We are willing to consider other projects in the field of structure-based drug discovery that match your interests and skills. Do not hesitate to get in contact.

Master project 2021-2022

Personal Information

Supervisor	Horacio Pérez-Sánchez
Email	hperez@ucam.edu
Institution	Universidad Católica de Murcia (UCAM)
Website	http://bio-hpc.eu
Group	Structural Bioinformatics and High Performance Computing Research Group

Project

Structural bioinformatics

Project Title:

Improvement and application of the blind docking technique

Keywords:

Docking, allosteric inhibitors, virtual screening, high performance computing

Summary:

Improvement and application of the Blind Docking technique developed in our group (https://bio- hpc.ucam.edu/achilles/). Several research lines are available in this context: a) application to drug discovery problems we are working on (Zika virus, colorectal cancer, diabetes, or others that you can propose), b) improvement of the blind docking algorithm, c) massive processing of the whole PDB database

References:

None

Expected skills::

None

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Chiara Pallara
Email	chiara.pallara@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Drug Discovery

Project

Structural bioinformatics

Project Title:

Frag-AA

Keywords:

Computational Chemistry, Molecular Modeling, python, Fragment Growing, Peptides

Summary:

We are currently developing packages to automate and standardize protocols oriented to use PELE, one of our proprietary softwares, for specific purposes in drug discovery. In this context, we have recently developed FragPELE, a new tool for in silico hit-to-lead drug design, capable of growing a fragment from a small molecule bound core while exploring the protein–ligand conformational space. This project will be focused on the development of FRAG-AA, a specific adaptation of FragPELE code to growing peptide-like molecules with the aim of developing protein-protein peptide-like inhibitors.

Expected skills::

Computer Sciences, Chemistry-Physics, Programming skills

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Initial duration: 6 months

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Baldo Oliva
Email	baldo.oliva@upf.edu
Institution	UPF
Website	http://sbi.upf.edu
Group	SBI

Project

Structural bioinformatics

Project Title:

Proposal to score protein structures using distance-restraints derived from multiple sequence alignments and statistic potentials.

Keywords:

Protein structure quality

Summary:

We work on the modeling of macro-complex regulatory structures formed by transcription factors (TF), co-factors, and DNA. One of the main contingencies to structurally model the macro-molecular complex formed by TFs and co-factors either in the enhancer or promoter binding sites is the completion of the structures. Modelling the structure of the DNA-binding domain of a TF bound to DNA can be achieved by homology modelling. However, the domains of TF that bind co-factors or other elements of the macro-molecular complex often involve unstructured or non-structurally solved regions. To solve this contingency we propose two approaches: 1) we will use docking of binary interactions if the structure of the unbound proteins is known; and 2) if the structure of one or both partners of a binary interaction we will apply threading and ab initio approaches to generate a structural model. Both approaches can be addressed by standard methods(1,2), but here we propose to update two innovations produced in our group during the previous grant. First, we developed iFrag to predict binding interfaces of proteins using only the sequences of the partners(3); second, we developed RADI (https://www.biorxiv.org/content/10.1101/406603v1), a program to predict residue-residue interactions based on amino-acid covariations in the sequence using Direct Coupling Analysis. We have used RADI, combined with local super-secondary structures, named sMotifs, as classified in ArchDB(4), to model ab initio the structure of proteins that cannot be modelled by homology modelling. We have already developed a pilot version, using restraints obtained with RADI, the prediction of secondary structure, and short fragment sMotifs templates from ArchDB to apply MODELLER(5) and constructed an ab initio model structure of a query sequence. However, we need to update this version to select the correct folds, because neither the restraints of RADI are exact, nor the short-fragment templates correct, and as a consequence, we obtain several but different solutions of the conformation. Therefore, we must cluster the solutions, score and rank the conformers and select the best conditions of the models. We have developed a program of statistical potentials, the SPserver (6) to select the correct folds. The scores of SPserver can be improved by training a combination of energies that benefits from a large knowledge of structures and the use of several potentials. The program can also benefit from the results of RADI to improve the learning. This machine-learning method will be useful to optimize the selection of structures. Proposed approach. 1) Define a method to compare as residue-profile the structures of model decoys with the experimental structure. This can be obtained by the superposition of the complete structure, using the RMSD fluctuation of CA atoms, the local matrix of distances around residues, or the local GDT score. Local distance matrices have the benefit to compare straightforward with potentials under a distance cut-off threshold. The information will be transformed into a score with values between 0 and 1, being 1 the best match (for example, the ratio of identic contacts within a radius that are the same between the decoy and the experimental structure). 2) Construct several ML methods of python “sklearn” library (SVM, NN, LogisticProgression, RF, etc.) using the statistic potential profiles per residue as inputs, the matrix of interacting potentials with the residues within a 3D-ball and the RADI direct and mutual information. The output will be a normalized score between 0 and 1 of the local quality. Additional approaches, using PyTorch, Keras and Tensorflow will also be considered. 3) The final score of the model will be obtained as the average of the total of scores along the sequence.

References:

1. Garcia-Garcia, J., Bonet, J., Guney, E., Fornes, O., Planas, J. and Oliva, B. (2012) Networks of ProteinProtein Interactions: From Uncertainty to Molecular Details. Mol Inform, 31, 342-362. 2. Mirela-Bota, P., Aguirre-Plans, J., Meseguer, A., Galletti, C., Segura, J., Planas-Iglesias, J., Garcia-Garcia, J., Guney, E., Oliva, B. and Fernandez-Fuentes, N. (2020) Galaxy InteractoMIX: An Integrated Computational Platform for the Study of Protein-Protein Interaction Data. J Mol Biol. 3. Garcia-Garcia, J., Valls-Comamala, V., Guney, E., Andreu, D., Munoz, F.J., Fernandez-Fuentes, N. and Oliva, B. (2017) iFrag: A Protein-Protein Interface Prediction Server Based on Sequence Fragments. J Mol Biol, 429, 382-389. 4. Bonet, J., Planas-Iglesias, J., Garcia-Garcia, J., Marin-Lopez, M.A., Fernandez-Fuentes, N. and Oliva, B. (2014) ArchDB 2014: structural classification of loops in proteins. Nucleic Acids Res, 42, D315-319. 5. Webb, B. and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics, 54, 5 6 1-5 6 37. 6. Aguirre-Plans, J., Meseguer, A., Molina-Fernández, R., Marin-Lopez, M.A., Jumde, G., Casanova, K., Bonet, J., Fornes, O., Fernandez-Fuentes, N. and Oliva, B. (2020) SPServer: Split-Statistical Potentials for the analysis of protein structures and protein-protein interactions. BMC Bioinformatics.

Expected skills::

Python programming.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Master project 2021-2022

Personal Information

Supervisor	Baldo Oliva
Email	baldo.oliva@upf.edu
Institution	UPF
Website	http://sbi.upf.edu
Group	SBI

Project

Structural bioinformatics

Project Title:

TF-DNA Binding strength affected by methylations

Keywords:

enhancer methylation

Summary:

Changes in DNA methylation are involved in development, disease, and the response to environmental conditions. Methylation of DNA is thought to regulate transcription both directly and indirectly. CpG methylation can directly repress transcription by preventing binding of some transcription factors (TFs) to their recognition motifs(17). For further insights, Lea et al. developed mSTARR-seq(18), a method that assesses the causal effects of DNA methylation on regulatory activity at genomic high-throughput level. Our objective is to predict the changes of TF binding caused by methylation. First we will build a database of methylated DNA binding with known TF binding. In a first approach the database will be extracted from experimental data of Yin et al. (17) and Lea et al. (18), indicating the loss or gain of TF binding. In a second approach, we will infer the effect from the comparison of bound TF binding sites with and without methylations. We will use the dataset of UniBind(13) to select the binding sites confirmed bound by TFs or the predictions of Viestra et al. (15). Then, we will select the tracks from UCSC Genome Browser with assays of DNA methylation (i.e. Methyl-RBBS) specific for tissue. We will compare the percentage that cytosines are methylated in the binding site with respect to any other location in the genome (this can be further refined by comparing with cytosines in the same TAD region). We will use the hypergeometric distribution to compare the ratio of methylation versus the expected ratio according to the length of binding recognition (as derived by the ChIP-Seq experiment). We will split the results in three categories: 1) If the ratio of methylation is lower than expected, then the methylation of cytosines reduces the TF binding. 2) On the contrary, if the methylation in the binding regions is higher than expected, the methylation is required for TF binding. 3) Otherwise, the methylation has no effect on TF binding. With the new bindings (case 2) we will generate statistical potentials specific of methylated cytosines by modelling the structure of TF-DNA binding, introducing a new symbol for methylated cytosines and including them in the general statistical potentials. We will calculate statistical potentials specific of the family of each TF including the new symbol for methyl-cytosine as in Meseguer et al. (19). The effect of disruption (case 1) will be used to generate statistical potentials specific of disruption. As before, the structure of TF-DNA binding will be modelled and the frequencies of the interactions between amino-acids and nucleotides will be obtained from the models. However, these potentials will be used to determine the potential of disruption, as these are the models of interactions lost after methylation. As before, a general potential will be derived with all TFs and their disrupted DNA binding sites and another set of potentials, specific for each TF family will be constructed. Finally, we will test the capacity of predicting TF disruptions after cytosine methylation or TF-DNA new bindings and specific PWMs for methyl cytosines. Two tests will be used for validation. First, using a 5-fold protocol with partially hidden data; and second, by training the method with one set of methylation (i.e. using the experiments of Yin et al. (17) and Lea et al. (18)) and testing the potentials in a different set (i.e. using data of ENCODE and removing redundancies with the training).

References:

1. Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. Nucleic Acids Res, 43, W39-49. 2. Fuxman Bass, J.I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D.E., Vidal, M. and Walhout, A.J. (2015) Human gene-centered transcription factor networks for enhancers and disease variants. Cell, 161, 661-673. 3. Fuxman Bass, J.I., Pons, C., Kozlowski, L., Reece-Hoyes, J.S., Shrestha, S., Holdorf, A.D., Mori, A., Myers, C.L. and Walhout, A.J. (2016) A gene-centered C. elegans protein-DNA interaction network provides a framework for functional predictions. Mol Syst Biol, 12, 884. 4. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. et al. (2013) DNA-binding specificities of human transcription factors. Cell, 152, 327-339. 5. Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Jr. and Reddy, T.E. (2015) Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res, 25, 1206-1214. 6. Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T.S. and Kellis, M. (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol, 34, 1180-1190. 7. Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D. and Shendure, J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol, 27, 1173-1175. 8. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B. et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively

parallel reporter assay. Nat Biotechnol, 30, 271-277. 9. Fornes, O., Gheorghe, M., Richmond, P.A., Arenillas, D.J., Wasserman, W.W. and Mathelier, A. (2018) MANTA2, update of the Mongo database for the analysis of transcription factor binding site alterations. Sci Data, 5, 180141. 10. Kumar, S., Ambrosini, G. and Bucher, P. (2017) SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res, 45, D139-D144. 11. Consortium, E.P., Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A. et al. (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature, 583, 699-710. 12. Wang, J., Zhuang, J., Iyer, S., Lin, X.Y., Greven, M.C., Kim, B.H., Moore, J., Pierce, B.G., Dong, X., Virgil, D. et al. (2013) Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res, 41, D171-176. 13. Gheorghe, M., Sandve, G.K., Khan, A., Cheneby, J., Ballester, B. and Mathelier, A. (2019) A map of direct TF-DNA interactions in the human genome. Nucleic Acids Res, 47, e21. 14. Mathelier, A., Shi, W. and Wasserman, W.W. (2015) Identification of altered cis-regulatory elements in human disease. Trends Genet, 31, 67-76. 15. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E. et al. (2020) Global reference mapping of human transcription factor footprints. Nature, 583, 729-736. 16. Meuleman, W., Muratov, A., Rynes, E., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Teodosiadis, A. et al. (2020) Index and biological spectrum of human DNase I hypersensitive sites. Nature, 584, 244-251. 17. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F. et al. (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science, 356. 18. Lea, A.J., Vockley, C.M., Johnston, R.A., Del Carpio, C.A., Barreiro, L.B., Reddy, T.E. and Tung, J. (2018) Genome-wide quantification of the effects of DNA methylation on human gene regulation. Elife, 7. 19. Meseguer, A., Årman, F., Fornes, O., Molina-Fernández, R., Bonet, J., Fernandez-Fuentes, N. and Oliva, B. (2020) On the prediction of DNA-binding preferences of C2H2-ZF domains using structural models: application on human CTCF. NAR Genomics and Bioinformatics, 2.

Expected skills::

Python programming

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Baldo Oliva
Email	baldo.oliva@upf.edu
Institution	UPF
Website	http://sbi.upf.edu
Group	SBI

Project

Structural bioinformatics

Project Title:

TF-DNA Binding strength affected by mutations on DNA binding site

Keywords:

Single nucleotide variants; Cis-regulation; Transcription factors;

Summary:

The principal reason to understand changes on the binding strength of a TF for a specific DNA binding sequence is to study cis-regulation and mutations that will affect it. Therefore, we focus on the prediction of the effect of one or more single nucleotide substitutions disrupting the recognition of the specific TF. The approach can be used to predict causal regulatory haplotypes that likely contribute to human phenotypes and to functionally fine map causal regulatory variants in regions of high linkage disequilibrium identified by expression quantitative trait loci (eQTL) analyses. We will predict binding strength to classify strong and weak changes and deciding if they imply the loss of the interaction. We will apply structural modelling with several templates, testing all potential PWMs, and the analysis of various statistical potentials on the TF-DNA interaction, comparing the native binding site with all other DNA variants. In order to train an AI/ML model, we will use the information from experimental Protein Binding Micro-arrays (PBM) of each TF. We will use as inputs the profiles of statisticalenergies and also the profiles of the PWM search along the DNA sequence, using the scores of FIMO (1) and the enrichment achieved with different structural models of the same TF. In order to learn and test with these scores, using different TFs and DNA binding lengths, the scores of the profiles will be normalized. The normalization will help us to better characterize the magnitude of the change produced by single nucleotide substitutions of the DNA binding sequence. The use of PBM will help us to handle an overwhelming amount of information, as all combinations of DNA sequences (formed by 8 or 12-mer nucleotides) are experimentally tested. We will train on PBM and the test will be performed on PBM (using a 10-fold approach), yeast-one-hybrid experiments on the specificity-changes of mutant DNA sequences(2,3), and other high-throughput experiments available such as SELEX(4), STARR-seq (5) or Sharpr-MPRA(6) (a modification of the MPRA(7,8) protocol that was developed to unveil at genome scale the effect of SNVs). Additional training and testing sets will be extracted from the datasets MANTA2 (9) and SNP2TFBS (10), composed by binding-site predictions in the human genome with the potential impact on TF binding for all possible SNVs. The impact of SNVs in MANTA2 is assessed by means of PWM scores computed on the alternate alleles. Similarly, we will use the theoretical PWMs plus all the energy profiles obtained by scanning the DNA sequence with the collection of TF-DNA structural models. Furthermore, the approach will be iteratively improved to be applicable on direct TF-DNA interactions in the human genome, using ChiP-seq data from recently expanded ENCODE encyclopaedia (11), updated versions of FactorBook (12), the dataset of UniBind (13) and datasets of altered cis-regulatory elements (14) and the location of DNase I hypersensitive regions of the genome with human genetic variation within transcription factor footprints(15,16)

References:

1. Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. Nucleic Acids Res, 43, W39-49. 2. Fuxman Bass, J.I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D.E., Vidal, M. and Walhout, A.J. (2015) Human gene-centered transcription factor networks for enhancers and disease variants. Cell, 161, 661-673. 3. Fuxman Bass, J.I., Pons, C., Kozlowski, L., Reece-Hoyes, J.S., Shrestha, S., Holdorf, A.D., Mori, A., Myers, C.L. and Walhout, A.J. (2016) A gene-centered C. elegans protein-DNA interaction network provides a framework for functional predictions. Mol Syst Biol, 12, 884. 4. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. et al. (2013) DNA-binding specificities of human transcription factors. Cell, 152, 327-339. 5. Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Jr. and Reddy, T.E. (2015) Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res, 25, 1206-1214. 6. Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T.S. and Kellis, M. (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol, 34, 1180-1190. 7. Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D. and Shendure, J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol, 27, 1173-1175. 8. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B. et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol, 30, 271-277. 9. Fornes, O., Gheorghe, M., Richmond, P.A., Arenillas, D.J., Wasserman, W.W. and Mathelier, A. (2018) MANTA2, update of the Mongo database for the analysis of transcription factor binding site alterations. Sci Data, 5, 180141. 10. Kumar, S., Ambrosini, G. and Bucher, P. (2017) SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. Nucleic Acids Res, 45, D139-D144. 11. Consortium, E.P., Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A. et al. (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature, 583, 699-710. 12. Wang, J., Zhuang, J., Iyer, S., Lin, X.Y., Greven, M.C., Kim, B.H., Moore, J., Pierce, B.G., Dong, X., Virgil, D. et al. (2013) Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res, 41, D171-176. 13. Gheorghe, M., Sandve, G.K., Khan, A., Cheneby, J., Ballester, B. and Mathelier, A. (2019) A map of direct TF-DNA interactions in the human genome. Nucleic Acids Res, 47, e21. 14. Mathelier, A., Shi, W. and Wasserman, W.W. (2015) Identification of altered cis-regulatory elements in human disease. Trends Genet, 31, 67-76. 15. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E. et al. (2020) Global reference mapping of human transcription factor footprints. Nature, 583, 729-736. 16. Meuleman, W., Muratov, A., Rynes, E., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Teodosiadis, A. et al. (2020) Index and biological spectrum of human DNase I hypersensitive sites. Nature, 584, 244-251. 17. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F. et al. (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science, 356. 18. Lea, A.J., Vockley, C.M., Johnston, R.A., Del Carpio, C.A., Barreiro, L.B., Reddy, T.E. and Tung, J. (2018) Genome-wide quantification of the effects of DNA methylation on human gene regulation. Elife, 7. 19. Meseguer, A., Årman, F., Fornes, O., Molina-Fernández, R., Bonet, J., Fernandez-Fuentes, N. and Oliva, B. (2020) On the prediction of DNA-binding preferences of C2H2-ZF domains using structural models: application on human CTCF. NAR Genomics and Bioinformatics, 2.

Expected skills::

Python programming

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Oriol Gallego
Email	oriol.gallego@upf.edu
Institution	UPF
Website	www.gallegolab.org
Group	Live-cell structural biology

Project

Structural bioinformatics

Project Title:

Integrative approaches to resolve protein structures in vivo

Keywords:

Integrative structural biology, Python, ImageJ, Pymol, CryoEM

Summary:

The mechanisms regulating biological processes are central questions in biomedicine. However, these mechanisms rely on protein structures and conformational dynamics that are often unknown. The complexity of the protein machinery involved, and fast cycles of assembly-activity-disassembly have prevented full understanding of the molecular basis that control human biology. For instance, in our group, we investigate the mechanism that drives exocytosis, a process that is essential in neurobiology and cell growth and whose mechanism of action is a long standing question in the field. We have developed a new method of fluorescent microscopy capable to resolve the 3D architecture of protein assemblies directly in living cells. Using this approach and computational integration of structural data we reconstructed de novo the exocytic machinery at the nanometre scale (Picco et al 2017 Cell). Our unpublished data indicates that the exocytic machinery organizes in transient higher-order structures with dimensions similar to the Nuclear Pore Complex, but with a highly dynamic behavior that prevents its purification and reconstitution in vitro. For these reasons, high-resolution structures and conformational dynamics necessary to understand the mechanism of exocytosis remain elusive. We offer a position for a Master student to continue the work published in Cell (Picco et al 2017 Cell). The student will use IMP (Integrative Modelling Platform, developed in A. Sali's lab), Python and image analysis to integrate in vitro and in cellulo datasets (i.e. live-cell imaging, cryo-EM, homology modelling, super resolution microscopy…) and to reconstruct the high-resolution structure of the supra-assembly that controls exocytosis. The project will be done in collaboration with the groups of Daniel Castaño (Biozentrum, Basel, Switzerland) and Alex de Marco (Monash University, Melbourne, Australia). The student is expected to contribute, together with experimentalists, in a larger project aiming to resolve the mechanism of exocytosis.

References:

Picco, A., Irastorza-Azcarate, I., Specht, T., Böke, D., Pazos, I., Rivier-Cordey, A-S., Devos, D.P., Kaksonen, M., Gallego, O., (2017) “The in vivo architecture of the exocyst provides structural basis for exocytosis.” Cell 168, 400-412.e18.

Expected skills::

Expertise with Python is required.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Baldo Oliva
Email	baldo.oliva@upf.edu
Institution	UPF
Website	sbi.upf.edu
Group	SBI

Project

Structural bioinformatics

Project Title:

Protein Folding ab initio, based on contact maps and supersecondary structures using distance-restraints derived from multiple sequence alignments.

Keywords:

ab initio fold prediction; protein design; threading

Summary:

Background We work on the modeling of macro-complex regulatory structures formed by transcription factors (TF), co-factors, and DNA. One of the main contingencies to structurally model the macro-molecular complex formed by TFs and co-factors either in the enhancer or promoter binding sites is the completion of the structures. Modelling the structure of the DNA-binding domain of a TF bound to DNA can be achieved by homology modelling. However, the domains of TF that bind co-factors or other elements of the macro-molecular complex often involve unstructured or non-structurally solved regions. To solve this contingency we propose two approaches: 1) we will use docking of binary interactions if the structure of the unbound proteins is known; and 2) if the structure of one or both partners of a binary interaction we will apply threading and ab initio approaches to generate a structural model. Both approaches can be addressed by standard methods(1,2), but here we propose to update two innovations produced in our group during the previous grant. First, we developed iFrag to predict binding interfaces of proteins using only the sequences of the partners(3); second, we developed RADI (https://www.biorxiv.org/content/10.1101/406603v1), a program to predict residue-residue interactions based on amino-acid covariations in the sequence using Direct Coupling Analysis. We plan to combine RADI and iFrag to improve the prediction of binding and apply the approach on directed docking, modelling the structure of binary interactions when the structures of both partners are known or modelled. Furthermore, we will use RADI, combined with local super-secondary structures, named sMotifs, as classified in ArchDB(4), to model ab initio the structure of proteins that cannot be modelled by homology modelling. Approach We have already developed a pilot version, using restraints obtained with RADI, the prediction of secondary structure, and short fragment sMotifs templates from ArchDB to apply MODELLER(5) and construct an ab initio model structure of a query sequence. However, we need to update this version to be applicable on a largest extend of proteins. First, we need to update the database of sMotifs. Second, many proteins can share similar local conformations but still they may not be classified. Therefore, we will extend the number of short-fragment templates by searching with psi-blast and selecting all short fragments with known structure that can potentially be used as templates. One of the problems already detected in the pilot is that more than one template can be assigned to the same region of the query sequence. We selected the longest fragment in the original version of the pilot, but the most appropriate approach is to model all combinations of fragments to produce several models after the application of restraints. Then, because neither the restraints of RADI are exact, nor the short-fragment templates correct, we will have several but different solutions of the conformation, so we will cluster the solutions, score and rank the conformers by statistical potentials using SPserver (6) and select the common restraints and common short fragment templates. The scores of SPserver can be improved by training a combination of energies that benefits from a large knowledge of structures and the use of several potentials. Also, we need a machine-learning method to optimize the selection of structures. A potential approach is to split the alignment in sections of 100 column-positions by sliding windows and construct the final model from the overlaps. The structure of the model will be iteratively repeated and improved until all convenient restraints and templates are congruent with the model.

References:

1. Garcia-Garcia, J., Bonet, J., Guney, E., Fornes, O., Planas, J. and Oliva, B. (2012) Networks of ProteinProtein Interactions: From Uncertainty to Molecular Details. Mol Inform, 31, 342-362. 2. Mirela-Bota, P., Aguirre-Plans, J., Meseguer, A., Galletti, C., Segura, J., Planas-Iglesias, J., Garcia-Garcia, J., Guney, E., Oliva, B. and Fernandez-Fuentes, N. (2020) Galaxy InteractoMIX: An Integrated Computational Platform for the Study of Protein-Protein Interaction Data. J Mol Biol. 3. Garcia-Garcia, J., Valls-Comamala, V., Guney, E., Andreu, D., Munoz, F.J., Fernandez-Fuentes, N. and Oliva, B. (2017) iFrag: A Protein-Protein Interface Prediction Server Based on Sequence Fragments. J Mol Biol, 429, 382-389. 4. Bonet, J., Planas-Iglesias, J., Garcia-Garcia, J., Marin-Lopez, M.A., Fernandez-Fuentes, N. and Oliva, B. (2014) ArchDB 2014: structural classification of loops in proteins. Nucleic Acids Res, 42, D315-319. 5. Webb, B. and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics, 54, 5 6 1-5 6 37. 6. Aguirre-Plans, J., Meseguer, A., Molina-Fernández, R., Marin-Lopez, M.A., Jumde, G., Casanova, K., Bonet, J., Fornes, O., Fernandez-Fuentes, N. and Oliva, B. (2020) SPServer: Split-Statistical Potentials for the analysis of protein structures and protein-protein interactions. BMC Bioinformatics.

Expected skills::

Python programming

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Master project 2021-2022

Personal Information

Supervisor	Toni Giorgino
Email	toni.giorgino@cnr.it
Institution	National Research Council of Italy
Website	www.giorginolab.it
Group	www.giorginolab.it

Project

Structural bioinformatics

Project Title:

Computational elucidation of large-scale activation dynamics in a multi domain protein

Keywords:

Multi-domain, state transition, metadynamics, conformational transition, gelsolin, amyloidosis, calcium.

Summary:

Gelsolin (GSN) is the prototype of a family of Ca2+ dependent proteins that regulate actin oligomerization state through their severing, capping and nucleating activities. Due to this polyvalent function and modular construction, gelsolin-like proteins are involved in several physiological processes. However, an exhaustive model of its physiologic mechanism of action is not yet available due to the highly dynamic structures of both gelsolin and actin. This project aims to develop a robust model of GSN activation by combining experimental and computational approaches. We will use extensive ("big data") molecular dynamics simulations with state-of-the art methodologies to elucidate the rare events and pathways underlying state transitions that are both physiologically relevant for the maintenance of the cell cytoskeleton, as well as in the pathology of diseases such as AGel amyloidosis. The work is essentially computational. The student will work with the direction of Dr. T. Giorgino [www.giorginolab.it] of the Institute of Biophysics of the Italian National Research Council. The work is part of an active scientific line conducted with the close collaboration of experimentalists (SAXS, crystallography, mutagenesis, stability assays). Due to the coronavirus situation, remote collaboration will be strongly preferred.

References:

- Giorgino T, Mattioni D, Hassan A, Milani M, Mastrangelo E, Barbiroli A, et al. Nanobody interaction unveils structure, dynamics and proteotoxicity of the Finnish-type amyloidogenic gelsolin variant. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2019 Mar 1;1865(3):648–60. - Chumnarnsilpa S, Robinson RC, Grimes JM, Leyrat C. Calcium-controlled conformational choreography in the N-terminal half of adseverin. Nat Commun. 2015 Nov;6(1):8254. - Nag S, Larsson M, Robinson RC, Burtnick LD. Gelsolin: The tail of a molecular gymnast. Cytoskeleton. 2013;70(7):360–84.

Expected skills::

The project is computationally oriented. Familiarity with Unix and interest in molecular simulations are essential. A degree of scientific independence is a plus.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Master project 2021-2022

Personal Information

Supervisor	Victor Guallar
Email	victor.guallar@bsc.es
Institution	Barcelona Supercomputing Center
Website	https://www.bsc.es/es/discover-bsc/organisation/scientific-structure/electronic-and-atomic-protein-modeling-eapm
Group	EAPM

Project

Structural bioinformatics

Project Title:

CovPELE

Keywords:

Drug design, covalent ligands, pharmacology, PELE, ligand binding, ligand diffusion

Summary:

We are seeking for a master student to develop and implement a new protocol for mapping covalent ligands diffusion and binding in pharmacological targets. The student will be part of a larger team and be responsible of: i) benchmark development, ii) protocol development, iii) protocol implementation (validation)

References:

A general view of PELE: Monte carlo techniques for drug design: the success case of PELE. JF Gilabert, D Lecina, J Estrada, V Guallar - Biomolecular Simulations in Structure-Based Drug Design, 2018 Adaptive simulations, towards interactive protein-ligand modeling. D Lecina, JF Gilabert, V Guallar. Scientific reports 7 (1), 1-11 (2017) An example of a recent implementation in PELE: FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. C Perez, D Soler, R Soliva, V Guallar. Journal of chemical information and modeling 60 (3), 1728-1736 (2020)

Expected skills::

Molecular modeling, python programming

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Master project 2021-2022

Personal Information

Supervisor	Javier Iglesias
Email	javier.iglesias@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Nucleic Acids

Project

Structural bioinformatics

Project Title:

Development of strategy for finding RNA ligands

Keywords:

Computational Chemistry, Molecular Modeling, Targeting RNA, Drug Design

Summary:

Targeting RNA is a new frontier in drug design, but a new molecular-level understanding of how to design agents for this is a critical need. RNA molecules are inherently flexible and, therefore, traditional rigid computational docking strategies are unsuited for the task. Therefore, strategies based on RNA conformational sampling studies followed by docking to the most populated states seem promising methodologies to develop.

Expected skills::

Biochemistry with programming skills or knowledge

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Initial duration: 6-8 months

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Lucía Díaz
Email	ldiaz@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Drug Discovery

Project

Structural bioinformatics

Project Title:

RNA-PROTACs

Keywords:

Computational Chemistry, Molecular Modeling, RNA Targeting, PROTACs, Drug Design

Summary:

PROTACs (PROteolysis-TArgeting Chimeras) are a new class of drugs that mediate protein degradation through the ubiquitination machinery. Conventional PROTACs comprise a ligand that binds to the target protein and a recruiting E3-ligase molecule that mediates protein degradation. Recently, RNA-PROTACs were introduced to facilitate the degradation of RNA binding proteins (RBPs), which are the origin of many diseases. In this project, a computational validation of already described RNA-PROTACs will be performed with the aim to develop a computational protocol for the design of novel molecules.

Expected skills::

Biochemistry with programming skills or knowledge

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

Initial project duration: 6-8 months

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Chiara Pallara
Email	chiara.pallara@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Drug Discovery

Project

Structural bioinformatics

Project Title:

In silico rational discovery of novel molecular glue degraders

Keywords:

Computational Chemistry, Molecular Modeling, Molecular Glues, Drug Design

Summary:

Molecular glue (MG) degraders induce protein–protein interactions that lead to ubiquitin-mediated protein degradation, catalysing the rapid depletion of previously inaccessible targets. This project consists in developing a computational protocol for the design of novel MG compounds, based on a Monte Carlo-based ligand landscape sampling within the possible protein-protein complex through the automatic combination of consecutive calls of pyDock and PELE, two of our proprietary softwares.

Expected skills::

Biochemistry with programming skills or knowledge

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

Initial project duration: 6-8 months

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Martí Municoy
Email	mmunicoy@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Software Development

Project

Structural bioinformatics

Project Title:

Development and testing a protocol for protein-peptide binding using PELE and BCE

Keywords:

Computational Chemistry, Molecular Modeling, Protein-Protein Interactions, QM/MM

Summary:

This project aims to combine PELE, our proprietary software to explore the potential energy surface of biomolecular systems, with the BCE workflow. BCE stands for Bioactive Conformational Ensemble and it combines a Molecular Mechanics exploration, a clustering technique and a final Quantum Mechanics calculation to obtain a conformational ensemble of drug-like molecules. The goal is to reduce the degrees of freedom to explore with PELE to only cover those conformations that BCE predicts as meaningful. This approach will allow us to study more efficiently protein binding of large peptides or other large molecules like macrocycles.

Expected skills::

Biochemistry with programming skills or knowledge

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Initial duration: 4-8 months

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Martí Municoy

Email	mmunicoy@nostrumbiodiscovery.com
Institution	NOSTRUM BIODISCOVERY
Website	http://nostrumbiodiscovery.com/index.html
Group	Software Development

Project

Structural bioinformatics

Project Title:

Development and validation of a ligand docking workflow for PELE

Keywords:

Computational Chemistry, Molecular Modeling, Docking,

Summary:

PELE is our Monte Carlo algorithm to explore the potential energy surface of biomolecular systems. Since its origin, 15 years ago, it has proved to be an outstanding tool to simulate migration pathways of ligands, predict binding modes and their affinities, or engineer enzymes. Recently, we have developed a platform to automate different workflows for drug design and enzyme engineering applications. However, we currently do not have a method to generate the initial structure with the ligand already inside the protein cavity that we want to study. As a consequence, we are forced to either predict the whole entrance of the ligand to the desired cavity, resulting in a very expensive calculation, or to dock it with external tools. The solution is to develop a method to generate this initial structure by trying to fit a 3D conformer of a small molecule into a specific protein cavity. Then, we could use some of the algorithms of PELE to sample and rank the different ligand binding modes. The goal of this project is to develop and validate a package to dock a ligand and rank its binding modes in a predefined protein cavity with PELE.

Expected skills::

Biochemistry with programming skills or knowledge

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Initial duration: 6-8 months