MASTER « *In Silico* Drug Design »

2ème année

PROPOSITION DE STAGE
Année Universitaire 2020/2021

A envoyer à Mme Pr Camproux
anne-claude.camproux@univ-paris-diderot.fr

Nom du Responsable du Laboratoire ou de l'Entreprise : Jean-Marie DUPRET

Affiliation administrative (CNRS, INSERM, …) et Numéro d'affiliation de l'unité : Université de Paris, Laboratoire "Biologie Fonctionnelle et Adaptative", CNRS UMR 8251, Equipe "Computational Modeling of Protein-Ligand Interactions", Inserm ERL U1133.

Adresse précise du Laboratoire : 35 rue Hélène Brion, 75205 Paris Cedex

Nom du Responsable de l'équipe d'accueil (EA) : Pierre TUFFERY
E-mail : pierre.tuffery@univ-paris-diderot.fr

Nom du Responsable du stage : Gautier MOROY & Guillaume POSTIC

Téléphone : 01 57 27 83 85 & 01 57 27 83 81
E-mail : gautier.moroy@univ-paris-diderot.fr & guillaume.postic@univ-paris-diderot.fr
HDR : oui

Ecole doctorale de rattachement : MTCI

Spécialité du stage : Recherche ☒ Professionnel ☐

Indiquez par quelques mots clés, l'orientation scientifique du sujet : développer une fonction de score spécifique aux interaction protéine-peptide et l'appliquer à l'identification de partenaires protéiques au peptide PIF.

Titre du stage : Identification of the Preimplantation Factor (PIF) molecular targets by virtual screening

Ce sujet constitue-t-il un premier pas vers un travail de thèse : ~~Oui - Non~~ Peut-être

---

Description du sujet (quelques lignes):

BACKGROUND

Preimplantation Factor (PIF) is a fifteen amino acid peptide, which is secreted during embryonic development in mammalian species, by trophoblast cells, prior to placenta formation. It plays essential roles in the initiation and maintenance of viable. Both *in vitro* and *in vivo* studies have identified several biological functions of PIF related to the embryo's neural system development and neuroprotection, as well as to autoimmunity. In animal models, the administration of a synthetic PIF can result in declined brain inflammation and increased neural repair.

These pathophysiological properties indicate a therapeutic potential of the PIF peptide for treating symptoms of Down syndrome (trisomy 21). The latter is one of the most common chromosome abnormalities in humans (1/1,000 babies born each year) and is notably associated with intellectual disability, as well as with various immune disorders.

OBJECTIVE

Despite its numerous biological functions, PIF's mechanisms of action remain unknown at the molecular level. This project aims (i) to identify the different protein partners of PIF *in silico* and (ii) to understand how PIF interacts with them to modify their biological activities.

METHODS

A virtual screening approach will be followed. For this purpose, we will use several 3D structures of the PIF peptide. These data are essential, as the biological activity of a peptide depends on its folding in space or conformation. At this stage, we have reliably predicted these structures by Replica Exchange Molecular Dynamics (REMD) simulations. By the end of this year, collaborators will provide experimental structural data, which will allow us to validate and refine the predicted 3D structures.

Task 1: virtual screening with classical bioinformatics tools

The Master intern will have to test the binding of PIF to the thousands of human proteins whose structures have been determined experimentally and deposited on the Protein Data Bank (PDB). For this purpose, s/he will apply a molecular docking protocol, called reverse docking, which consists in searching the compatibility between a molecular ligand and its potential receptor, regarding different criteria (electrostatic interaction, hydrophobicity, van der Waals interaction, hydrogen bond, ...). This will be a high-throughput process, as it will require ~100,000 docking simulations, using the computational resources of our laboratory (RPBS platform).

Task 2: development of a PIF-specific computational method

The Master intern will also accomplish a more fundamental task, in terms of bioinformatics development. Indeed, virtual screening is commonly used in the pharmaceutical industry to eliminate the worst molecular candidates, thus reducing the costs of *in vitro* screening. Thus, bioinformatics programs for docking have already been developed and made available. For a given ligand, these tools are able to explore the surface of a protein receptor for searching the different potential binding sites. This first step of a docking procedure is called “sampling” and is carried out efficiently by any available tool. However, this is not the case for the other docking step called “scoring”. The latter consists in predicting the ranking of the best docking “poses”: the best 3D structural models produced for a given peptide-protein complex. The available scoring functions display limited success, as they rarely correlate with experimental free energy of binding. Moreover, these available scoring functions have been optimized for small molecules and are not adapted to the specificities of protein-peptide interactions.

Thus, the task for the Master intern will be to develop such a scoring function. This will be the direct application of our fundamental and recently published work in the field of knowledge-based potentials (DOI: 10.1016/j.csbj.2020.08.013). Our open-source computational tool for building user-defined scoring functions works as a Bayesian classifier. It has shown high accuracy for the intra-protein atomic interactions, but has not been evaluated for inter-protein atomic interactions. Therefore, the Master intern will have to build a training dataset from the recent database PepPro (DOI: 10.1002/jcc.26114), which is a non-redundant set of experimental structures of protein-peptide complexes. Regarding the molecular docking, these experimentally determined poses will be considered as the “ground truth”. The Master intern will generate incorrect poses purposely, by applying the aforementioned docking protocol for each ligand-receptor pair of the PepPro database. The quality of these false poses will be measured by their distance to the corresponding true pose. The generated dataset (~15,000 receptor-ligand poses) will then be used to train a statistical model. The resulting scoring function will be benchmarked on this same dataset, following a cross-validation procedure. The performance obtained will be compared to that of other scoring functions from the literature. Finally, the new scoring function will be applied to re-rank the poses generated by the virtual screening and, thus, identify the best candidate protein targets of the PIF peptide.

Retour par e-mail : anne-claude.camproux@univ-paris-diderot.fr