



MODELO DE OFERTA DE PRÁCTICAS CURRICULARES

Título de las prácticas:

Título atractivo, con las palabras clave de la actividad o temática de las prácticas, o su ubicación.

Machine-learning scoring functions for structure-based virtual screening

Descripción de las funciones del alumno

Molecular docking is a family of computational tools frequently used in drug design. Docking aims at predicting whether and how a small molecule binds to a macromolecular target. Classical scoring functions (SFs), based on linear regression, have reached a plateau in their performance to carry out such prediction. By contrast, machine-learning SFs have been widely praised by the degree with which they have outperformed classical SFs at binding affinity prediction. Research on the optimal application of Machine Learning (ML) to the related problem of Structure-Based Virtual Screening (SBVS) is highly promising, but it is still in its infancy.

References

1. Fresnais, L., Ballester, P.J. (2020) "The impact of compound library size on the performance of scoring functions for structure-based virtual screening". Briefings in Bioinformatics, bbaa095 (jointly with a 2019 M2 student). <https://doi.org/10.1093/bib/bbaa095>
2. Li, H., Sze, K-H., Lu, G., Ballester, P.J. (2020) "Machine-learning scoring functions for structure-based virtual screening". WIREs Computational Molecular Science, e1478. <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1478>
3. Li, H., Sze, K-H., Lu, G., Ballester, P.J. (2020) "Machine-learning scoring functions for structure-based drug lead optimization". WIREs Computational Molecular Science, e1467. <https://onlinelibrary.wiley.com/doi/10.1002/wcms.1465>
4. Hongjian Li, Jiangjun Peng, Yee Leung, Kwong-Sak Leung, Man-Hon Wong, Gang Lu, Ballester, P.J. (2019) "Classical scoring functions for docking are unable to exploit large volumes of structural and interaction data". Bioinformatics 35, 3989-95. <http://dx.doi.org/10.1093/bioinformatics/btz183>

Requisitos: (indicar titulación y curso); otros requisitos adicionales (idiomas, informática, otros conocimientos, etc).

The ideal candidate will be:

- Skilled in the implementation of python (numPy/SciPy/scikit-learn) scripts for scientific data analysis.
- Previously exposed to supervised learning algorithms.
- Familiar with open-source cheminformatics toolkits (e.g. RDKit) and medicinal chemistry databases (e.g. ChEMBL).
- Trained in structural bioinformatics and/or computational drug design

Proyecto formativo

El objetivo fundamental de la Práctica Externa es guiar al alumno para que aplique en el mundo



real sus conocimientos, destrezas y habilidades, en un entorno de trabajo en grupo, que reproduzca las condiciones que se pueden encontrar en su futuro lugar de trabajo. Las funciones y tareas a desarrollar en la Práctica permitirán ayudar al alumno a desarrollar sus competencias profesionales desde tres dimensiones: competencias técnicas (conocimientos técnicos propios de la titulación); competencias personales (comportamientos, comunicación, sentido de responsabilidad, compromiso y motivación, creatividad e iniciativa, implicación, trabajo en equipo) y competencias contextuales (capacidad de adaptación al contexto profesional)

Módulo TRABAJO FIN DE GRADO. El objetivo fundamental del TFG es la realización de un trabajo académico que demuestre que el alumno es capaz de aplicar los conocimientos y competencias que ha adquirido a lo largo de la carrera para tratar de resolver un problema, aprovechar una oportunidad o satisfacer una necesidad, de similar naturaleza y complejidad a los que pueda desarrollar en el ejercicio de su actividad profesional, eligiendo una solución que sea viable, tanto desde un punto de vista técnico como económico.

Actividades a desarrollar en la práctica académica:

We are looking for a student to carry out her/his 6-month M2 project in this area. More concretely, we are interested in investigating the performance of ML-based SFs for SBVS of targets for which few ligands are known by transfer learning. We have already developed code with several ML algorithms and featurisation schemes, so the student will be able to concentrate on running and analysing the numerical experiments rather than having to generate this infrastructure first. The other main task of the student will be to document in depth all the research steps so that the protocols are available for others to reproduce and build upon this research.

Nº de plazas:	1
Fecha de inicio:	February-2021
Fecha de fin:	August-2021
Horas semanales:	35h
Horario jornada laboral:	
Importe Ayuda/Bolsa de estudio:	500€/mes
Tutor académico:	
Email:	



Departamento tutor académico:	
Tutor empresa:	Dr Pedro Ballester
Email tutor empresa:	pedro.ballester@inserm.fr
Departamento tutor empresa:	Cancer Research Center of Marseille https://www.crcm-marseille.fr/
ENTIDAD COLABORADORA:	INSERM
<i>A cumplimentar por Oficina Prácticas:</i> Créditos a reconocer (Nº ECTS):	

Enviar por email a: Secretaria.pei.etsiaab@upm.es
(Susana Pardo – Tfno: 91 06 70764)