POLITÉCNICA

E.T.S. DE INGENIERÍA AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

Título de las prácticas (Title of the internship):

Large-scale annotation of protein residues based on domain structures

Descripción de las funciones del alumno (Description of the student´s tasks)

Advances in high-throughput sequencing have led to an unprecedented growth in genome and protein sequences being submitted to biological databases. UniProt is a world-leading resource providing open access to a database of protein sequences and function. UniProt develops computational methods to annotate millions of protein sequences for thousands of species. This project aims to research the use of protein structure information for computational annotation of proteins at large-scale. For this, the student will investigate Funfams (Functional families) which provides further clustering to CATH (a protein structure classification database) based on the protein functions. Each FunFam provides a multiple sequence alignment of the protein sequences within the cluster. The alignment allows to transfer functional site annotations from PDB from one protein to all other proteins within the same cluster. The student's tasks will be:

- Select a dataset of proteins within the FunFam clusters, that contain site-annotations of interest from PDBe-KB
- Select a diverse set of 15-20 FunFam clusters and transfer annotations collected in the previous step to additional proteins
- Verify the plausibility of the transferred annotations (e.g. through available 3D-Structures or 3D-models)
- Expand the procedure to all FunFam clusters
- Develop a prototype for annotation of protein residues in UniProt
- At the end of the project the below questions should be answered:
  - How many functional site annotations from PDBe-KB can be transferred to how many new proteins?
  - How many FunFam clusters can be used to transfer site annotations from PDBe-KB?
  - Are there any conflicts between different site annotations within a FunFam cluster?
  - How FunFam can be used for functional annotation of protein residues?

Requisitos (Prerequisites): *(indicar titulación y curso) (give Grade and academic year); otros requisitos adicionales (idiomas, informática, otros conocimientos, etc) (other aditional prerequitistes (languages, informatics, other knowledge, etc)*

Student in the Grade of Masters. The student should have a good background in one programming language e.g. Phython. Interest in data science and some biological background would be helpful. The student should be proficient in the English language.

Proyecto formativo (Training Project)

Module EXTERNAL PRACTICES. The fundamental goal of the external practices is to guide the student in applying his previously adquired knowledge to real tasks in a group work environment the realistically represents the work conditions the the students will encounter in their future roles. In this way, the student will be able to get familiar with a working environment (work schedule, responsability, attitude, organization, etc),and with the adequate working methodology in profesional reality, contrasting and applying the acquired academic knowledge.

Actividades a desarrollar en la práctica académica (Activities that will be performed in the academic internship):

The student will learn about biological databases, their content and use. S(he) will get familiar with relational databases, SQL, and programming supported by a team of software engineers and bioinformaticians. The student will perfom data science tasks e.g. query database content, use Jupiter notebooks and/or other statistical tools to represent data, and develop software to analyse data. The student will participate in group meetings, and present his/her work to the team. The student will have the opportunity to meet key researchers in the bioinformatics field.

Nº de plazas: **(Nr. of places)**	**1**
¿El alumno tendrá trato habitual	**No**

con menores? **(Has the student dealings with underage persons?)**	
Fecha de inicio: **(Starting date)**	**Any time in 2022**
Fecha de fin: **(End date)**	**4-6 months**
Horas semanales: **(Weekly hours)**	**25-39**
Horario jornada laboral: **(Working hours)**	**8h**
Importe Ayuda/Bolsa de estudio: **(Amount of fellowship / remuneration)**	**€/mes**
Tutor académico: (Academic tutor (UPM)) Email: martin@ebi.ac.uk	**Dr. Maria J. Martin**
Departamento tutor académico: **(Dept. of academic tutor)**	**Protein Function development, EMBL-EBI, UK**
Tutor empresa:	

(External tutor)	
Email tutor empresa: **(Email external tutor)**	
Departamento tutor empresa: **(Dept. of external tutor)**	
Ubicación de la estancia de las practicas **(Location of the internship)**	
ENTIDAD COLABORADORA: **(Collaborating Entity)**	
A cumplimentar por Oficina Prácticas ETSIAAB: **Créditos a reconocer (Nº ECTS)**:	

Enviar por email a: OFICINA DE PRÁCTICAS ACADEMICAS EXTERNAS – ETSIAAB
secretaria.pei.etsiaab@upm.es – Secretarias: Visitación Pérez / Susana Pardo - Tfno: 913363686)