PhD thesis proposal within IMGT, Montpellier, FRANCE

Combining IMGT/2D/3Dstructure-DB with AI methodologies to unravel the secrets of paratope & epitope complementarity

IMGT®, the international ImMunoGeneTics information system® http://www.imgt., is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR) and major histocompatibility (MH) of human and other vertebrate species. IMGT® is the current reference in immunogenetics and immunoinformatics [1]. It was created in 1989, more than 30 years ago, by Marie-Paule Lefranc at Montpellier, France.

With the increase in availability of 3D structures from early 2000s, IMGT/2Dstructure-DB and IMGT/3Dstructure-DB [2] were developed in 2001, for immunological and related molecules. Data computationally curated under expert supervision on daily basis feed these bases which currently (March 2020) account 6800 entries in total and approximately 550 new entries in IMGT/3Dstructure-DB and 150 new entries in IMGT/2Dstructure-DB are added every year.

Objectives

Year 1

The student will propose a solution and develop a prototype based on the IMGT/2D/3Dstructure-DB workflow to implement a novel, optimized and robust pipeline that will help to analyze upcoming data with more complexity and re-analyze old data regularly and in a synchronized manner with the continuously updated IMGT germline database. The student will contribute to the feeding of the databases in order to have a clear understanding 1) of the IMGT biocuration process, 2) of the different types of data stored in these databases 3) as well as of the different external sources. These data will be used in the 2nd year. The student will write a bioinformatics article on IMGT/2D/3Dstructure-DB and the novel features in them.

Year 2

The student will focus on paratope & epitope and/or contact analysis data and will implement different deep learning methodologies in order to explore whether a paratope, the part of the antibody which recognizes and binds to the antigen, can be predicted from a given eptitope and which are the important amino acids in the contact between a paratope and an epitope. The student will have to scrutinize the scientific literature in the field, such as [3-5] and have a good understanding of the state of the art of the field before implementing his/her methodologies.

Year 3

The student will add functionalities to the already available code and/or develop a software package with his/her code and make it available via the IMGT portal. The student will write up his/her thesis and a 2nd publication which will be describing the chosen methodology and the software package developed.

Risk management / Significance in the field

Being a computational project, the risks are minimal. The biological question of the prediction of the epitope is a long-standing challenge and a panacea might not be found, at the end of the PhD thesis, however the explorative work will certainly enlighten our understanding. The importance of the scientific question is unquestionable. Accurately predicting the paratope of an antibody given the epitope, opens up a highway to the treatment of any existing and/or emerging immunology based disease.

Skills required

Master degree or equivalent in bioinformatics. Software engineering or similar discipline. English verbal and writing skills. Good team player, interest in transdisciplinary field.

Starting date between 1st of September 2020 and 1st of December 2020.

For more information, contact: sofia.kossida@igh.cnrs.fr

References

1. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G, Kossida S. IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. 2015 Jan;43(Database issue):D413-22. doi: 10.1093/nar/gku1056. Epub 2014 Nov 5.PMID: 25378316
2. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF.Ehrenmann, F., Kaas, Q. and Lefranc, M.-P. Nucl. Acids Res., 38, D301-307 (2010). Epub 2009 Nov 9; doi:10.1093/nar/gkp946.PMID: 19900967
3. Antibody Complementarity Determining Region Design Using High-Capacity Machine Learning. Liu G, Zeng H, Mueller J, Carter B, Wang Z, Schilz J, Horny G, Birnbaum ME, Ewert S, Gifford DK. Bioinformatics. 2019 Nov 28. pii: btz895. doi: 10.1093/bioinformatics/btz895. [Epub ahead of print] PMID:31778140
4. Progress and challenges in predicting protein interfaces. Esmaielbeiki R, Krawczyk K, Knapp B, Nebel JC, Deane CM. Brief Bioinform. 2016 Jan;17(1):117-31. doi: 10.1093/bib/bbv027. Epub 2015 May 13. Review.
5. Rationalization and design of the complementarity determining region sequences in an antibody-antigen recognition interface. Yu CM, Peng HP, Chen IC, Lee YC, Chen JB, Tsai KC, Chen CT, Chang JY, Yang EW, Hsu PC, Jian JW, Hsu HJ, Chang HJ, Hsu WL, Huang KF, Ma AC, Yang AS. PLoS One. 2012;7(3):e33340. doi: 10.1371/journal.pone.0033340. Epub 2012 Mar 22. PMID:22457753