MARIE SKŁODOWSKA-CURIE INDIVIDUAL FELLOWSHIPS 2019

EXPRESSION OF INTEREST FOR HOSTING MARIE CURIE FELLOWS

HOST INSTITUTION

NOVA School of Science and Technology | LAQV – Associated Laboratory for Green Chemistry

RESEARCH GROUP AND URL

Chemoinformatics Lab
http://sites.fct.unl.pt/quimioinformatica

SUPERVISOR (NAME AND E-MAIL)

João Aires de Sousa
jas@fct.unl.pt

SHORT CV OF THE SUPERVISOR

João Aires de Sousa was born in Lisbon, Portugal, in 1970. He studied chemistry at the New University of Lisbon (FCT/UNL), where he graduated in 1993, and obtained his Ph.D. (1997) in Organic Chemistry with A. Lobo and S. Prabhakar. His Ph.D. thesis was on 'Chiral Synthesis of N-Arylaziridines'. In 1998 he joined the group of J. Gasteiger in Erlangen, Germany, as a post-doctoral fellow, and in 2002 he was appointed as an Assistant Professor of the Department of Chemistry at the New University of Lisbon (Portugal). In 2010 JAS obtained the Habilitation in Chemistry at FCT-UNL.

JAS leads a research group of Chemoinformatics since 2002. He has coordinated 15 postdoc researchers and has supervised 4 PhD students. His main scientific interests include automatic learning of chemical reactivity and metabolism, genome-scale classification of enzymatic reactions, prediction of NMR spectra, and representation of molecular chirality for computer prediction of observable properties.

João Aires de Sousa has designed and implemented Chemoinformatics modules at undergraduate and graduate levels at FCT-UNL (Caparica, Portugal) and FC-UL (Lisbon, Portugal). He has taught Chemoinformatics at the University of Sao Paulo (Brazil) and regularly teaches at the Master of Chemoinformatics of the University of Strasbourg (France).

JAS is engaged in the implementation of active learning methodologies, namely Team-Based Learning, in a range of curricular units, from General Chemistry to Organic Chemistry to Chemoinformatics. He has coordinated the Applied Chemistry Degree ("Licenciatura em Química Aplicada", FCT-UNL, 2009-2011) and the Bioorganic Master course at FCT-UNL (2011).

JAS is currently a research partner of the Shikifactory100 H2020 project, a member of the international Rabbit doctoral training programme and the Portuguese coordinator of a PESSOA joint research project with the University of Strasbourg.

ORCID: http://orcid.org/0000-0002-5887-2966

5 SELECTED PUBLICATIONS

- Pereira F, Aires-de-Sousa, J., "Machine learning for the prediction of molecular dipole moments obtained by density functional theory", *J. Cheminform.* **2018**, *10*:43.

- Pereira F, Xiao K, Latino DA, Wu C, Zhang Q, Aires-de-Sousa, J., "Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals", *J. Chem. Inf. Model.* **2017**, *57*(1), 11-21.

- Zhang Q, Zheng F, Zhao T, Qu X, Aires-de-Sousa J, "Machine Learning Estimation of Atom Condensed Fukui Functions", *Molecular Informatics* **2016**, *35*, 62–69.

- Q.-Y. Zhang, F. Zheng, R.P.S. Fartaria, D. A. R. S. Latino, X. Qu, T. Campos, T. Zhao, J. Aires-de-Sousa, "A QSPR approach for the fast estimation of DFT/NBO partial atomic charges", *Chemom. Intell. Lab. Syst.* **2014,** *134*, 158-163.

- X. Qu, D.A.R.S. Latino, J. Aires-de-Sousa, "A big data approach to the ultra-fast prediction of DFT-calculated bond energies", *J. Cheminform.* **2013**, *5*(1), 34.

PROJECT TITLE AND SHORT DESCRIPTION

Machine learning for the fast estimation of quantum chemical reactivity properties

The rapid access to properties of molecules concerning their chemical reactivity is highly desired for large-scale data mining explorations, e.g., in drug discovery projects requiring early toxicity risk assessment, or in the development of new materials. Large volumes of data are being produced by quantum chemistry calculations, which provide increasingly accurate estimations of several properties, e.g. by Density Functional Theory (DFT), but are still too computationally expensive for large scale uses. This project explores the possibility of using large amounts of data generated by DFT methods for thousands of molecular structures, extracting relevant molecular properties and applying machine learning (ML) algorithms to learn from the data. Once trained, these ML models can be applied to new structures to produce ultra-fast predictions.

The project will involve the generation of databases of DFT calculations for thousands of bond energies focusing on specific molecular substructures. Beyond such local reactivity properties, databases of global molecular parameters related to ionisation will also be obtained, and explored with ML algorithms. The final models will be incorporated in libraries that can be imported by chemoinformatic software packages.

SCIENTIFIC AREA WHERE THE PROJECT FITS BEST

Chemistry (CHE)