Marie Skłodowska Curie Action – Postdoctoral Fellowship 2023
Expression of interest – Hosting offer
(MSCA-PF-2023)

Contact Person/Scientist in charge *(data of the principal investigator of the research group/lab or scientific supervisor)* **Name**	Thijs
Surname	Stuyver
Email	thijs.stuyver@chimieparistech.psl.eu
Department /Institute /Centre *(data of the centre/department where the fellow would be located* **Name**	CTM group / Institute of Chemistry for Life and Health Sciences (iCLeHS)/ Chimie ParisTech PSL
Address	11 rue Pierre et Marie Curie 75005 Paris
Research Area *(Please select the research area: corresponding to the eight MSCA evaluation panels. You can select between one and up to three scientific areas per EOI)*	**Physics (PHY)** **Chemistry (CHE)**
Brief description of the Centre/Research Group/Team *(max. 1,600 characters including spaces: information about the research centre or research group, scientific staff. Please include URL if possible)*	This project will be conducted at the Institute for Life and Health Sciences (i-CLeHS), directed by Prof. Carlo Amado, of Chimie Paristech – PSL University, in the "*Chemical Theory and Modelling*" (CTM) group. The research group is composed of 5 permanent staff members, 2 post-docs and 6 PhD students. Several research lines related to electronic structure method development and artificial intelligence are being developed within the group, and ample computational resources are available both on-site and at the national level. More specifically, over 1600 CPU cores are available locally, and the group also has a generous allocation on the national "*Grand equipment national de calcul intensif*" (GENCI) supercomputer. Thijs Stuyver is Junior Professor at Chimie Paristech since January 2023, and his research can be situated at the interface between computational chemistry and data science (https://thijsstuyver.com).
Project description / Topic of interest *(max. 1,800 characters including spaces: short description of the research project / research line where the fellow would be hosted and develop his /her project)*	Traditional machine learning algorithms tend to require lots of training data. However, by introducing domain knowledge, one can dramatically improve the data efficiency -- as well as generalizability -- of these models. In this research line, we aim to augment models with descriptors stemming from quantum chemical calculations to facilitate their application to domains for which only limited data is available. An available project within the machine learning team of the research group consists of investigating the connection between the roughness of molecular property landscapes and the inclusion of domain knowledge inspired descriptors. The roughness/smoothness of molecular property landscapes, i.e., the evolution of the target property as one traverses the selected feature space, has often been connected to the inherent modellability of the target property. Featurizations that give rise to smooth activity landscapes ought to result in machine learning models requiring relatively few data points, whereas rough landscapes ought to result in models requiring many data points to reach an acceptable accuracy. In this project, we would first investigate how the inclusion of physics inspired descriptors affects the roughness of activity landscapes, and subsequently we would

	determine whether changes in roughness can be used as an unbiased metric to select informative descriptors for a machine learning model. Other projects related to computational chemistry and data science can also be discussed; please reach out through email (with your CV and a letter of motivation attached) for more information.
Applications: documents to be submitted and deadlines	CV, letter of motivation Deadline: end of April 2023.