POLITECNICO DI MILANO

PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 39th cycle

Research Area n. 1 - Computer Science and Engineering

PNRR 118 PNRR Research Field: HIGH-THROUGHPUT VIRTUAL SCREENING TARGETING EXASCALE HPC INFRASTRUCTURES AND URGENT COMPUTING

Monthly net income of PhDscholarship (max 36 months)
€ 1400.0 In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity	
Motivation and objectives of the research in this field	The recent pandemic has shown the need to develop and test drugs in a very short time. High Throughput Virtual Screening (HTVS) is an in silico phase that enables scientists to test thousands of compounds (drug candidates) simultaneously using virtual models. So far most of the investments in the field were not in computational tools, but rather in wet chemistry, and clinical trials. In recent years, the situation changed and HPC (High-Performance Computing) has been used in the field of drug discovery to accelerate the process of finding new drugs since it permits researchers to analyze large chemical libraries against selected targets in a short amount of time, allowing them to identify potential treatments more quickly, and anticipating the possibility of failures in the wet phases. In this context, molecular docking is the most widely used computational method in structure-based drug design when the 3D structure of the molecules (both ligand and target) is known. It relies on the shape complementarity search of a flexible ligand, adopting a rigid target model and a quick interaction evaluation using empirical scoring functions. Despite this method permits a fast screening, the molecule ranking suffers from low accuracy in terms of interaction prediction, which usually leads to a low hit rate in subsequent experimental validations. To overcome this

	problem, molecular simulation methods targeting binding free energy prediction between a receptor and its ligand are used (e.g., FEP-ABFE). However, these methods are computationally intensive and cannot be used to screen many compounds in a reasonable time. Starting from the successful use of HPC techniques in the context of an extreme-scale molecular docking platform, the goal of the research will be to analyze the problems of adopting a hybrid approach involving both Molecular docking and Molecular Modeling techniques. This will be done by enhancing approximate high-throughput molecular docking approaches with biochemical features of the compounds and by approximate molecular modeling calculations to keep under control the computational complexity of the evaluations, as needed by the urgent computing scenario, considering the target screening goal that is not an accurate energy estimation. Hybrid approaches efficiently using the underlying large-scale and heterogeneous (GPU-based) HPC infrastructures are currently not yet available in the literature. The success of the research will be measured by considering different orthogonal KPIs: - Efficient use of the resources, i.e. number of resources needed to complete a Virtual Screening (VS) campaign in a fixed amount of time - Performance improvement, i.e. time needed to complete a VS campaign on a fixed amount of resources - Screening accuracy, i.e. the capability of identifying active targets given a fixed amount of compute hours (resources and time)
Methods and techniques that will be developed and used to carry out the research	The research is based on multidisciplinary aspects that span from the annotation of biochemical features inherited by accurate molecular simulations to the acceleration of Molecular Modeling techniques using HPC architectures and approximation approaches. The novelty of the research is the possibility to codesign a hybrid molecular docking/modeling technique capable of efficiently exploiting modern HPC infrastructures for large-scale virtual-screening campaigns. The three main directions

	that will be addressed by the student are the following: 1) Lightweight annotation of the ligand and binding site to improve docking accuracy while keeping performance. Molecular modeling techniques (e.g. using force fields derived by AMBER and CHARM tools) will be contextualized in the high-throughput urgent scenario deriving the trade-off accuracy performance. 2) Approximation of Molecular Modeling techniques for large-scale virtual screening on HPC architectures. E.g. Identification of unnecessary calculations to be performed to trade off screening accuracy with performance in the context of molecular simulation and binding free energy estimation; Efficient porting of the defined workflow on the heterogeneous and large-scale infrastructure. 3) Use of machine learning techniques to reduce the computational effort needed for the virtual screening of huge databases of ligands. E.g. the results derived from the previous studies can be used to support recommender systems or more in general machine learning algorithms in the context of drug discovery.
Educational objectives	The candidate will be trained and pushed to work towards techniques related to the efficient use of HPC resources and approximate computing in the context of drug discovery. Different approaches for GPU programming and performance portability, as well as methods to navigate the accuracy-performance trade-off curve, will be part of the training for the student. The urgent computing scenario that is the background of this interdisciplinary research, creates a unique possibility for a PhD in terms of novel methods, techniques, and experiments, targeting both the supercomputing and the molecular modeling field.
Job opportunities	The job opportunities in the field are several since research work covers the aspects from both HPC and Drug Discovery fields, and thus the wide areas of computational science and GPU programming. Possibilities are among academic, supercomputing/research centers (e.g. CINECA, CSC) or industrial positions (e.g. Dompè Farmaceutici, NVidia, Leonardo).

Composition of the research group	1 Full Professors 1 Associated Professors 3 Assistant Professors 2 PhD Students
Name of the research directors	prof. Gianluca Palermo

Contacts
gianluca.palermo@polimi.it - +39 022399 3552 - https://tinyurl.com/webGianluca23

Additional support - Financial aid per PhD student per year (gross amount)	
Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
Amount monthly	700.0 €
By number of months	6

National Operational Program for Research and Innovation	
Company where the candidate will attend the stage (name and brief description)	VSB - Technical University of Ostrava - IT4Innovation Czech Republic Supercomputing Center
By number of months at the company	6
Institution or company where the candidate will spend the period abroad (name and brief description)	VSB - Technical University of Ostrava - IT4Innovation Czech Republic Supercomputing Center
By number of months abroad	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information
EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student. TEACHING ASSISTANTSHIP: availability of funding in recognition of supporting teaching activities by the PhD student There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations. COMPUTER AVAILABILITY: individual use. DESK AVAILABILITY: individual use.


POLITECNICO
MILANO