2020 Helmholtz – OCPC – Program
for the involvement of postdocs in bilateral collaboration projects

PART A

Title of the project: Large scale protein simulation techniques and structure data mining

Helmholtz Centre and institute: Forschungszentrum Jülich, Jülich Supercomputer Centre (JSC), Institute for Advanced Simulation (IAS)

Project leader: Dr. Olav Zimmermann (Simulation Laboratory Biology, JSC, Jülich)

Web-address:
http://www.fz-juelich.de/SharedDocs/Personen/IAS/JSC/EN/staff/zimmermann_o.html
http://www.fz-juelich.de/ias/jsc/EN/AboutUs/Organisation/ComputationalScience/Simlabs/slbio/_node.html

Description of the project:
Markov Chain Monte Carlo (MCMC) simulations enable the study of biomolecular phenomena that act on longer time scale by generating useable statistics at a small fraction of the computational cost of comparable Molecular Dynamics (MD) simulations. In contrast to MD simulations, MCMC can already provide novel insights into large-scale peptide aggregation and other long-time scale phenomena in atomic detail and is therefore an important method for sampling the molecular conformations from different potential energy regions. MCMC simulations constitute a form of importance sampling, and can generate billions of sample conformations and Tera-Bytes of data. The tools available for analysis and interpretation of large volumes of simulation data tend to be geared towards MD rather than MCMC data, and lack the scope and parallel scalability required to cope with the sampling capacity of MCMC simulations.

Previous project: The current joint postdoc project is a continuation of a previous project funded by Helmholtz and OCPC. Dr. Runyu Jing, who was funded by the phase 1 joint postdoc program, has performed well and now works as assistant professor at Sichuan University, Chengdu, China. Within the MCMC-structure analysis framework MCMine Dr. Jing developed parallel methods for analyzing the trajectories generated from MCMC simulations on the Tera-Byte scale. This included parallel analysis modules implemented in C++ for fundamental patterns of molecular structures, a set of descriptors to characterize peptide aggegation, and a library of python scripts for downstream analysis and visualization. He successfully applied this framework to several extreme scale data sets from aggregation simulations of the medically relevant peptide PHF6 to detect and characterize patterns in the formation of large peptide clusters.

Project aim: Building on this research the current project aims at extending the application spectrum of MCMine to cover other large scale structure formation processes in particular protein folding and protein interaction. To this end new descriptors will need to be developed and tested.

Tests will be performed on the supercomputers at the National Supercomputing Center in Shenzhen, China, and at the Forschungszentrum Jülich. It is also intended to demonstrate that such high performance analysis tools can play a key role in the detection of artifactual bias in current molecular force fields and will, in combination with Machine Learning algorithms, enable the development of much improved simulation models for bio-macromolecules.

Description of existing or sought Chinese collaboration partner institute:
The SimLab Biology at JSC would like to extend its fruitful collaboration with the:
Computational Biology and Bioinformatics Group at Shenzhen Institutes of Advanced Technology (SIAT).
Head: Prof. Yanjie Wei, Ph.D., Professor, Email: yj.wei@siat.ac.cn, Phone: +86-755-86392336
Web: http://hpcc.siat.ac.cn/cbb-lab/
The research areas of Prof. Wei's group at SIAT's Center for High Performance Computing include protein folding and structure prediction; gene assembly, classification and clustering. The group has developed several open source projects in these areas including a massively parallel genome assembler: http://sourceforge.net/projects/swapassembler/, a novel clustering algorithm for genes: https://github.com/Huiyang520/Dmk-Bkmeans, and a hybrid algorithm for residue-contact prediction of trans-membrane proteins: http://hpcc.siat.ac.cn/COMSAT/.

The group has a perfectly fitting profile with first grade experience in high performance computing, protein structure prediction and machine learning. Prof. Wei was a frequent visitor in Juelich, as his PhD supervisor at Michigan Tech University, Prof. Uli Hansmann was also the head of a NIC research group at Juelich, the members of which are now the core team of the Simulation Laboratory Biology at JSC.

With currently 2000 staff Shenzhen Institutes of Advanced Technology is a fast growing member of the renowned Chinese Academy of Sciences. It is comprised of nine institutes and numerous other labs and facilities, many of them with IT or biomedical focus.

Required qualification of the post-doc:

- PhD in Bioinformatics, (Bio-)physics, Computer Science or related
- Experience with Programming required (C++ and Python preferred)
- Additional skills in Molecular Simulation, Statistics, Software Engineering and experience in working in a Linux Environment are a plus

PART B

Documents to be provided by the post-doc, necessary for an application to OCPC via a postdoc-station in China, which is affiliated to a research institution like a university:

- Detailed description of the interest in joining the project (motivation letter)
- Curriculum vitae, copies of degrees
- List of publications
- 2 letters of recommendation
- Proof of command of English language

PART C

Additional requirements to be fulfilled by the post-doc:

- Max. age of 35 years
- PhD degree not older than 5 years
- Very good command of the English language
- Strong ability to work independently and in a team