Enclosure 1b. Category 1 Application form – English version

APPLICATIONS ARE PREFERABLY DRAWN UP IN ENGLISH. AN ENGLISH TRANSLATION HAS TO BE ENCLOSED WITH APPLICATIONS SUBMITTED IN DUTCH.

The application form is available in English on the website

https://www.vscentrum.be/en/access-and-infrastructure/project-access-tier1

Title of the application:

Study of oxygen diffusion through cell membranes

Name and first name of the applicant:

De Vos, Oriana

Institution:

Ghent University

Research group / department:

Center for Molecular Modeling (CMM)

Title / position:

Assistant

e-mail address:

Oriana.DeVos@UGent.be

Total computing time that is needed, in node days:

1125 nodedays

Total disk storage that is applied for (in GiB):

540 GB

1. Title of the research project (with IWETO or FRIS link if available) within the framework of which computing time is applied for:

Molecular simulations on the role of (dis)order in rafts on oxygen diffusion through cell membranes

2. Describe your research project in short. Explicitly mention the scientific questions that you are planning to address and the overall scientific goals of the project. (max. 1 A4 in Arial 12):

Everywhere in the human body, you can find membranes. Around every cell and organelle, there is a membrane. Membranes consist of phospholipid bilayers. Phospholipids are composed of a hydrophilic head and two hydrophobic tails. In the bilayer of membranes the hydrophobic tails are facing each other.

In membranes, there can be found ordered domains rich in cholesterol, sphingomyelin (a group of phospholipids) and saturated phospholipids. These nanoscale domains, called rafts, are surrounded by disordered domains. The rafts function together with membrane proteins, both integral and peripheral proteins. There are indications that rafts play a role in signal transduction and intracellular transport. Cholesterol is known to have an influence on the flexibility of the cell membrane.

Oxygen must be transported through the cell membrane. It is important for the energy supply to the cell. The central question of this project is “how can oxygen diffuse through the membrane?”. Here specifically, this project will investigate whether rafts have an influence on the oxygen transport. We expect that the transport might be easier in rafts than in disordered domains. In rafts the pathway of oxygen would be (almost) straight, but in disordered domains the pathway would be longer and would contain more curves. Oxygen mobility will be examined using computational modeling.

First, the project will focus on the correct description of oxygen interactions. It will be investigated whether the standard force field parameters provide accurate dynamical information, i.e. whether viscosity and diffusion constants may be reproduced. Second, the oxygen diffusion through rafts-containing membranes will be modelled in follow-up research. The post-processing of the molecular dynamics trajectories will provide us with detailed information on the oxygen pathways through the cell membranes.

3. Provide an abstract (10 lines) for scientific communication on the website in layman's terms. See also item 12 of this application form.

   Everywhere in the human body, there are membranes, for instance around every cell and organelle. Oxygen must be transported through the cell membrane in order to supply energy to the cell. The central question of this project is "how can oxygen diffuse through the membrane?" This project will investigate whether the presence of ordered rafts in the membrane have an influence on oxygen transport. We expect that the transport might be easier in ordered rafts than in disordered domains. Oxygen mobility will be examined using molecular dynamics simulations, where the initial step is the correct description of oxygen interactions at the atomic level.

4. Financing institution or channel, financing the research project in full or in part (FWO, BOF, IWT, EU, ...): Please attach the confirmation letter as enclosure (see instructions in enclosure 3 "EasyChair proposals submission procedure").

   Oriana De Vos is an assistant at Ghent University. (This round of TIER1-proposals is without charge, so we are not foreseeing financing for this project.)

5. Name and email address of the promoter(s) of the research project:

   Prof. An Ghysels: An.Ghysels@UGent.be
   Prof. Tanja Van Hecke: Tanja.VanHecke@UGent.be

6. Persons mandated by the Applicant to compute on the Tier-1 within the framework of the present project: Please provide for every person:
   - name and first name
   - institution
   - research group / department
   - title / position
   - experience of using HPC resources in the past (Tier-0/Tier-1/Tier-2 infrastructure in Belgium and abroad)

   Oriana De Vos (vsc41441)
   Center for Molecular Modeling, Ghent University
   Assistant
   - Has started using the HPC infrastructure since April 2015 (TIER2).
   - Has developed experience with CHARMM after a successful application of a TIER1 Starting Grant (Oct 2015).

Prof. An Ghysels (vsc40051)
Center for Molecular Modeling, Ghent University
Has experience with a variety of computational codes, such as CHARMM in collaboration with foreign groups.
Has experience since 2007 using HPC infrastructure.

- on Biowulf (National Institutes of Health (NIH, USA), global cluster),
- on LoBoS (Lots of Boxes on Shelves, Laboratory of Computational Biology, NIH, USA)
- on VSC (UGent): early user of gengar/TIER2, has previously been awarded a TIER1 project (2013)

7. Explain why this project needs to run on a Tier-1 system, why the machine you have requested is suitable for the project and how the use of the system will enable the science proposed (max. ½ A4 in Arial 12).

   This project consists of one type of task: molecular dynamics simulations, where atomic trajectories are created by integrating Newton's equation. The challenge is here the size of the system (over 10,000 atoms) and the amount of simulations needed (200 nanoseconds per membrane to get statistics). Such **long** simulations of **large systems** need to be modelled using parallellized software to make the calculations feasible in an acceptable amount of wall time.
   The task is well defined: we will run molecular dynamics simulations of the membrane where we vary only a few parameters: with different oxygen interaction models and different membrane compositions. Each new system has about the same size and requires 200 nanoseconds of simulation time (about 2e8 integration time steps). Running many similar long simulations seems most suitable to be done on the TIER1 infrastructure, and can be seen as **high-throughput usage of the cluster**. Our tests in the Starting Grant (see Table 1/Plot 1) show that the CHARMM software scales very well on TIER1, demonstrating that TIER1 is suitable hardware for the simulations.

8. Justify the number of node days requested. This should include information such as: number and nature of computing tasks, software used, and the sequence in which they will be performed.
   Indicate for each typical computing task the required resources:

   - wall clock time (note that 3 days is the maximal wall clock time for any job; checkpointing should be used for longer run times)
     We will use 3 days as wall clock time. We plan on breaking up the long time-scale trajectories into smaller trajectories. This allows for better management of the output (post-processing) and restarting simulations after the 3 days wall clock time. Restart files

will be written out about every hour. This gives a good trade-off between safety (restart) and performance (limiting i/o).

- memory (maximum 64 GiB/node)
  We will run the simulations on full nodes and the estimated memory use of a computing task is well below 64 GiB/node. The memory per node is therefore largely sufficient.

- number of nodes: see Table 2
- number of CPU cores: see Table 2
- disk space (estimated volume in GiB and the total number of files); make a clear distinction between usage of Tier-2 DATA/HOME partitions and the Tier-1 SCRATCH partition :
  see Table 2 / see below
- number of tasks, and an indication of how many such tasks would be submitted concurrently. See Table 2 / see below.

This information should take the form of a table (an example is provided as Table 2 in the appendix). Provide additional descriptions of the computing tasks and comments as needed. Resource estimates should be preferably based on the results of actual calculations on Tier-1 (via, e.g., a Starting Grant) for system/problem sizes that are on par with those of the intended computing tasks (e.g., same mesh sizes, actual molecular system, …). If not, provide the name, architecture, #cores, memory, etc. of the machine that was used to obtain these results and explain how you have calculated/rescaled the wall clock times, number of cores, etc.

(max. 1 A4 Arial 12).

+ see Table 2

We have one particular computing task: running molecular dynamics simulations with the modelling package CHARMM.
The TIER1 Starting Grant was used to test on a model system of equal size.

We define a "run" as a molecular dynamics simulation of 1 nanosecond, taking 10^6 time steps of 1 femtosecond each. The equilibration time of the volume of the simulation box takes 50 runs (50 nanoseconds). Subsequently, data should be collected in the so-called production time. We have noticed in previous work (collaboration with NIH) that 200 runs are necessary to obtain sufficient statistics. This corresponds to 0.2 microseconds of data, i.e. 2e8 time steps, and this time scale is long enough to observe oxygen diffusion. This gives **250 runs** in total per modelled membrane system.

One run takes 9 hours on 1 node on the TIER1 cluster according to our tests in the Starting Grant. As each system takes 250 runs, it takes 2250 hours, which equals 250*9/24 = **94 node days**.
We now count the number of modelled systems. Four membranes will be simulated with various phospholipids, cholesterol, and cardiolipin: (1) POPC, (2) DOPC, (3) DPPC, (4) mitochondrial composition. Three oxygen models will be simulated: (1) standard force field parameters, (2) larger van der Waals radius for oxygen, (3) quadrupole model for oxygen. The latter interaction model will only be marginally more expensive (effect is limited to 10 oxygens in the systems). This brings the total number to **12 membrane systems**.
With 94 nodedays per membrane system, we thus arrive at a total of 12*94 = **1125 nodedays** for this project.

The total number of tasks is thus 12*250 = **3000 tasks**.
At most 12 simulations would be submitted concurrently, one for each system.
Disk space: one run produces about **180 MB of data per run**, as coordinates are saved every 1000 time steps. With 250 runs per system, we arrive at 250*180MB = 45 GB of data per system. For 12 membrane systems, we arrive at 12*45 = **540 GB** for the total of this project.
During each run, only 180 MB should be stored on the TIER1-SCRATCH. This 180 MB of data can than be moved towards the TIER2-HOME directory after each run. This means that scratch storage should be no issue for our molecular dynamics simulations.
To perform a run, CHARMM only needs about **15 files** (input files, restart files) on TIER1. Each run creates 4 new files. The 3000 tasks will therefore generate 3000*4 = **12 000 files**, which may be moved to the TIER2-HOME.

9. Describe the software required to perform the computing task(s). Please clearly provide the following per item in this regard:

   - a reference to the software's web page
   - the software license system (open source, GPL, etc.)
   - if there is no free academic use of the software, state which license makes the installation and the use valid on the Tier-1 by the Applicant (+ add a copy of the signed license)
   - if need be, which license server will be used (name + IP address)
   - whether the software is already available on the Tier-1 (see https://www.vscentrum.be/cluster-doc/software/tier1-muk) and, if this is not the case, compilation and installation instructions (possibly with reference to existing Tier-2 installation)

CHARMM

- http://www.charmm.org/
- CHARMM has a commercial licence system. We have used the license previously in the TIER1 Starting Grant. After purchase the license is valid forever, while a new license should be purchased for new releases of the program.
- The software is already available on the TIER1 infrastructure.

Provide the results of scaling tests that were conducted with this software, preferably on Tier-1 (using, e.g., a Starting Grant) for system/problem sizes that are on par with those of the intended computing tasks (e.g., same mesh sizes, actual molecular system, …). If not, provide the name, architecture, #cores, memory, etc. of the machine that was used to obtain these results.
Provide both a table and scaling plot such as table 1 and plot 1 in the appendix (max. 2 A4 in Arial 12).

See Table 1 and Plot 1. The testing system had equal size as the membrane systems in this project.

10. Describe how you will manage the resources requested in the period during which the task is to be performed. What usage pattern do you anticipate (similar usage on monthly basis, bursts, …)?

    After setting up the simulations (which might take two weeks), we expect similar continuous usage per month throughout the project period. Intermediately, the simulations will be checked in order to verify if the simulations are proceeding correctly.

11. List the granted computing time allocations to the promoter(s) of this research project, on the Flemish Tier-1 system, as well as other Tier-1 and Tier-0 systems. Also, describe the scientific output obtained within the framework of computing time that was granted during the past two years on the Flemish Tier-1 or on other Tier-1 or Tier-0 supercomputers. DOI links are sufficient.

    TIER1 grant (An Ghysels, Samuel Moors)
    Diffusion of hydrocarbons in porous frameworks
    4320 nodedays, 07/07/2014 – 31/12/2014
    Output: http://dx.doi.org/10.1021/acs.jpcc.5b06010

    TIER1 Starting Grant (Oriana De Vos, promotor An Ghysels)

100 nodedays, 02/10/2015-31/01/2016
Study of the oxygen diffusion through cell membranes
Output: /

12. Are the applicants of this application bound by a confidentiality agreement? If so, the title and the abstract of this application will not be published on the website of the FWO / Flemish Supercomputer Center.

No.

Should you have any questions or encounter any difficulties during the electronic submission of an Application, please contact by e-mail:
Associatie KU Leuven: hpcinfo@kuleuven.be
Associatie Universiteit Gent: hpc@ugent.be
Associatie Universiteit Hogescholen Antwerpen: hpc@uantwerpen.be
Associatie Universiteit Hogescholen Limburg: geertjan.bex@uhasselt.be
Universitaire Associatie Brussel: rosette.vandenbroucke@vub.ac.be
For the other institutions: caroline.volckaert@FWO.be

Appendix: tables and plots

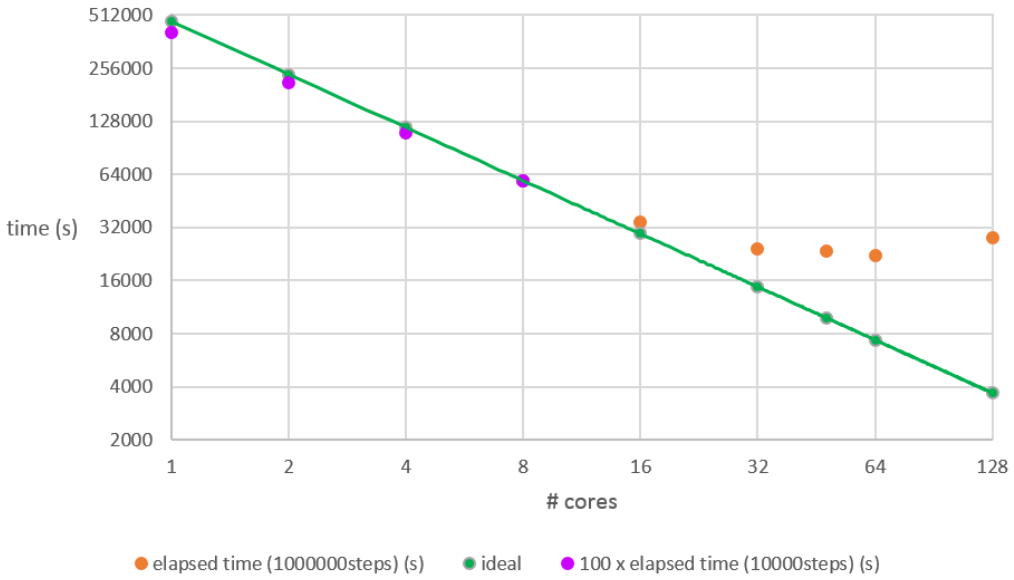
Table 1

# nodes	# cores	absolute timing (s)	speedup	# cores x timing
1	1	410400	1.0000	410400
1	2	211620	1.9393	423240
1	4	110460	3.7154	441840
1	8	59160	6.9371	473280
1	16	34128 *	12.0253	546048
2	32	24264 *	16.9139	776448
3	48	23544 *	17.4312	1130112
4	64	22140 *	18.5366	1416960
8	128	27864 *	14.7287	3566592

Remarks:

- CPU times are almost equal to elapsed times.
- Absolute times labelled with a star (*) are computed with 1000000 time steps in the molecular dynamics. Other CPU times, without a star, are computed with 10000 time steps and have been multiplied by 100 to allow for comparison.

Plot 1



Plot 1 Elapsed time of 10^6 time steps as a function of processors: elapsed time computed with 10^6 time steps (orange), elapsed time computed with 10^4 time steps and multiplied by 100 (purple), and the expected time assuming ideal scaling (green).

Table 2

	Node day calculation							Storage volume estimate	
Computational task	# of such tasks	Wall clock time (days) per task	# Tier-1 nodes per task	# node days per task	# CPU cores per task	Memory usage (GiB) / node per task	OpenMP / MPI / hybrid / vSMP	Tier-2 DATA/HOME volume (GiB) + number of files	Tier-1 SCRATCH volume (GiB) + number of files
Task1	3000	0.375	1	1125	16	< 64 GB	MPI	540 GB 12 000 files	0.18 GB 15 files