Enclosure 1. Tier-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:

In silico study of lymphoma-related MyD88 mutations and their effect on protein actions and mechanism.

Name and first name of the applicant:

Peelman Frank

Institution:

Ghent University, VIB

Research group / department:

VIB-UGent Center for Medical Biotechnology

Title / position:

Full Professor

e-mail address:

frank.peelman@vib-ugent.be

Total computing time that is needed, in node days:

1676.8 node days

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

15.36 GB Tier-1 SCRATCH, 76.8 GB Tier-2 DATA

The total number of pages in this application should not exceed 17, excluding possible appendices (confirmation letter of financing institution,

software license,…) which may be taken into account by the Tier-1 Allocation Board.

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

Analyse van signalisatiecomplexen aan de hand van hedendaagse interactomica

http://www.researchportal.be/project/analyse-van-signalisatiecomplexen-aan-de-hand-van-hedendaagse-interactomica-(UG_6287070700)/

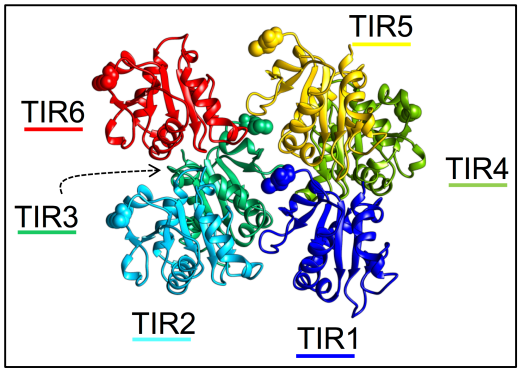
2. Describe your research project in short. Explicitly mention the scientific questions that you are planning to address and the overall scientific goals of the project. (max. 1 A4 in Arial 12):

The human protein MyD88 is a crucial intracellular adaptor protein for signal transduction of the Toll-like receptors (TLRs) and the receptors of the interleukin-1 family. These receptors are essential for our innate immunity against invading pathogens and for the initiation of a proper adaptive immunity.

It is now clear that MyD88 is a cancer driver gene, as several different MyD88 missense mutations are found in various B-cell lymphoma, e.g. in Waldenström's macroglobulinemia (WM) patients, MALT or Burkitt's lymphoma. These mutations are all found in the MyD88 TIR domain, which the protein uses to homo-oligomerize and to interact with the TIR domain of TLRs or interleukin-1 family receptors.

While the structure of the MyD88 TIR domain has been determined, it was not clear how these domains interact. In previous studies, we have characterized these interactions, relying on a host of biochemical experiments such as extensive random mutagenesis, augmented with microsecond molecular dynamics simulations (Vyncke et al. 2016). Via this combined approach, for instance, we have demonstrated that the L265P mutation (associated with WM) in the core of the MyD88 TIR domain leads to profound structural rearrangement in one of the binding sites, facilitating MyD88 TIR-TIR interactions.

We now want to understand how all other MyD88 mutations in B-cell lymphoma types affect MyD88 activation and its interactions. Already, the necessary tools to rapidly analyze *in cyto* the effects of the mutations on the signaling and interactions of MyD88 have been set up. To link this data to a molecular mechanism, we want to perform microsecond molecular dynamics analysis for eight MyD88 mutants. This will help to understand how lymphoma-related MyD88 mutations affect the actions and mechanism of MyD88. We hope to understand whether specific mutations are linked to different disease development processes. This can possibly support prognosis and therapeutic decision making.



References

"Reconstructing the TIR Side of the Myddosome: a Paradigm for TIR-TIR Interactions", Vyncke et al., Structure. 2016 Mar 1;24(3):437-47.

3. Provide an engaging abstract (10 lines) for scientific communication on the website in layman's terms. Should this application be bound by a confidentiality agreement (see also item 12 of this application form), provide more details about the specific nature of the confidentiality and indicate why an abstract may not be published.

The protein MyD88 provides a crucial signaling function for several receptors that are essential for human immunity against invading pathogens. In various cancer types, it has been found that part of this protein has undergone mutation. In this project, we examine several mutations in this protein that lead to various types of white blood cell cancers. Via biochemical experiments, we analyze the effects of these mutations on the protein function. With the aid of microsecond molecular dynamics, we link this experimental data to molecular mechanisms. This helps us understand how these mutations affect the actions and mechanism of MyD88, and how they are connected to disease development processes.

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. In case the project has not gone through a scientific approval process attach a letter of approval of your own institute.

IUAP-VII (1/1/2012-31/12/2017)

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

Prof. Jan Tavernier (jan.tavernier@ugent.vib.be)

Prof. Frank Peelman (frank.peelman@ugent.vib.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)

- Prof. Dr. Peelman Frank
  - Ghent University, VIB
  - VIB-UGent Center for Medical Biotechnology
  - Prof. Peelman has +10 years of expertise in various molecular modelling techniques on a dedicated Tier-2 cluster.
- Dr. Pauwels Ewald
  - Ghent University
  - ICT department
  - This project taps into expertise that Dr. Pauwels developed during research work prior to his current role as UGent HPC coordinator. He will guide and help with the technical execution of all simulations.
  - Dr. Pauwels has over 13 years of experience in using the UGent Tier-2 research infrastructure, and has experience in running Tier-1 simulations (as pilot user), both on MUK and BrENIAC.

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).

Considerable computational resources are required for this project, which are scarcely available on Tier-2 systems. Running all simulations there would result in long waiting times in the queue, incurring a massive delay to the project.

We would like to use the Tier-1 to speed up the execution of the production MD runs for all eight virtual mutation experiments considered in this project. These would be executed early on in the new Tier-1 allocation, in the july-august 2017 timeframe, so that computational results can augment the experimental analysis that is nearly finished and will be published in the PhD thesis of L. Vyncke (VIB-UGent Center for Medical Biotechnology) in October 2017. All equilibration runs will be executed on Tier-2, ahead of a potential Tier-1 allocation.

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;)
   - memory (maximum 128 GiB/node; 256 GiB/node is available upon motivated request)
   - number of nodes
   - number of CPU cores
   - 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
   - number of tasks, and an indication of how many such tasks would be submitted concurrently.

   This information should take the form of a table (an example is provided as Table 1 on the next page). Provide additional descriptions of the computing tasks and comments as needed and clearly relate the described tasks to the tasks in the table. 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.

   (typically up to 2 A4 Arial 12).

In this project, we will run extended molecular dynamics simulations with GROMACS on eight mutated versions of the MyD88 structure. We will rely on a computational protocol (Vyncke et al. 2016) that, for each mutant, consists of the following MD runs (time step 1 fs):

- 5 x 100 ns equilibration
- 10 x 100 ns production runs for 4 replicas (with randomized velocities)

For ease of use and in order to constantly monitor sanity of the molecular dynamics, simulations are split up in runs of 100 ns. All equilibration runs will be prepared on Tier-2 clusters, leaving for Tier-1:

8 (mutations) x 4 (replicas) x 10 (100-ns runs) = **320** 100-ns runs

Each 100-ns run takes 1.31 days on 4 nodes, bringing the total requested compute time to 1676.8 nodedays. The rationale for choosing 4 nodes per run (instead of e.g. 2 nodes) is that this offers the best throughput-time, while maintaining a reasonable efficiency (see section 9).

Table 1

	Node day calculation							Storage volume estimate	
Computational task	# of such tasks	Wall clock time (days) per task	# Tier-1 nodes per task	# total node days task	# CPU cores per task	Memory usage (GiB) / node per task	OpenMP / MPI / hybrid	Tier-2 DATA/HOME volume (GiB) + number of files	Tier-1 SCRATCH volume (GiB) + number of files
100-ns production runs for 8 mutations, 4 replicas	8 x 4 x 10	1.31	4	(8 x 4 x 10) x 1.31 x 4	112	49 GB / node	Hybrid	8 x 4 x 10 x 240 Mb 8 x 4 x 10 x 15 files	8 x 4 x 2 x 240 Mb 8 x 4 x 2 x 15 files
Total	320 tasks			1676.8 node days				76.8 Gb, 4800 files	15.36 Gb, 960 files

In this table, task2 needs to run 5 times. One run of task2 uses 10 nodes for 2 days and it uses all cores in the node. In this case the node has 28 cores, so the task2 uses 280 cores. The task needs 64GB RAM in each node. The task needs 100 GB of scratch disk space. If you plan to run the tasks concurrently mention this in the description, so you can specify a correct total number of scratch space that you need.

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 and, if this is not the case, compilation and installation instructions (possibly with reference to existing Tier-2 installation)

   Provide the results of scaling tests that were conducted with this software, preferably on the current VSC 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 run on the current VSC Tier-1, provide the name, architecture, #cores, memory, etc. of the machine that was used to obtain these results and how you think this compares to the current VSC Tier-1. If a different system/problem size is used provide some guidance how it relates to the problem size in the application.
   Provide both a table and scaling plot such as table 2 and plot 1 below (typically up to 3 A4 in Arial 12).

All simulations will be performed with GROMACS v2016.3 (http://www.gromacs.org). This is free software, available under the GNU Lesser General Public License (LGPL), version 2.1.

Version 2016.3 of GROMACS is already installed on BrENIAC and this Tier1 cluster has been used for all scaling results. Scaling tests were performed by running a 100-ps MD simulation on one of the mutated MyD88 proteins, i.e. on one of the actual biochemical systems that will be examined in this project. Results are given in Table 2. We didn't examine parallel usage beyond 8 nodes, as the speedup already starts to level off at this point (see Plot 1). For 8 nodes, efficiency drops below 70% (Plot 2).

Table 2

# nodes	# cores	Wall clock time (s)	Speedup (with respect to 1 node)	Efficiency
1	28	318.927	1.00	1.000
2	56	170.836	1.86	0.933
4	112	93.542	3.41	0.852
8	224	57.553	5.54	0.693

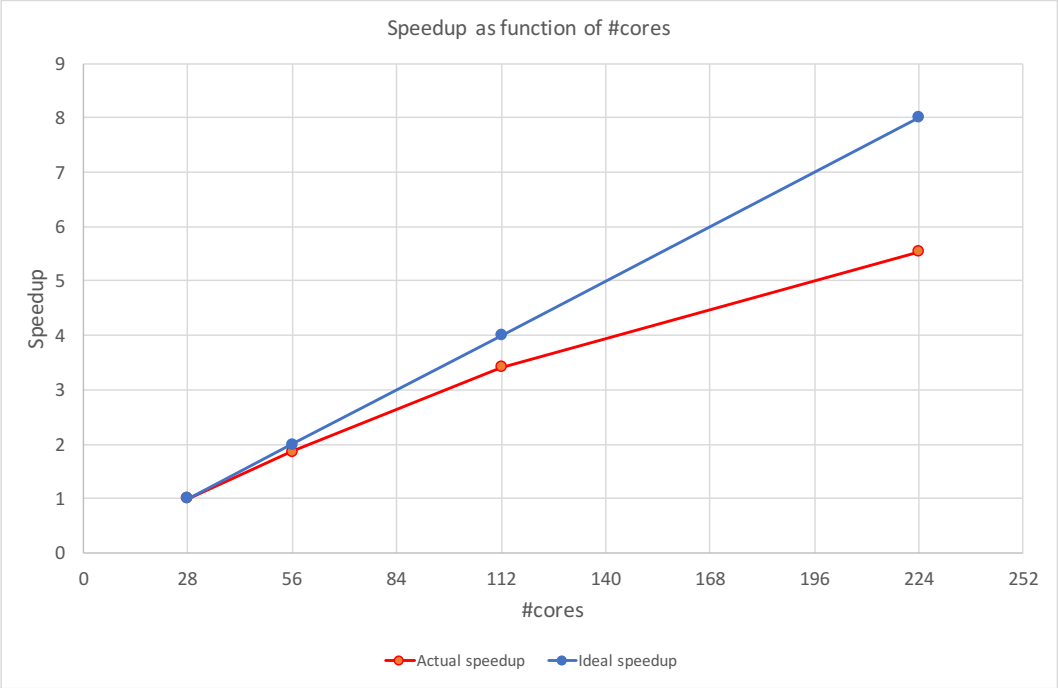
Three further experiments were performed on the same system, for 1000x longer 100-ns MD runtimes. Results are shown in Table 3. (No such simulation on 1 node was performed, since it could not fit within the 3-day wallclock limit on BrENIAC. As such, speedup and efficiency in Table 3 are calculated based on a hypothetical 102.42 hour wallclock for a 1-node simulation. This is under the assumption that the performance for going from 100-ps to 100-ns MD runs would be the same as for the 2-node case: 170.8 s / 100 ps = 1.71 s wallclock / ps of simulation → 54.86 h / 100 ns = 1.98 s wallclock / ps of simulation; factor = 1.156.)

In view of both sets of scaling experiments, we choose to run all MD simulations using 4 nodes. This maintains efficiency at 82%-85% but offers a better throughput-time for the entire set of 10 x 100-ns MD runs per replica: ideally 10 x 31.23 h = 13 days wallclock instead of 23 days wallclock.

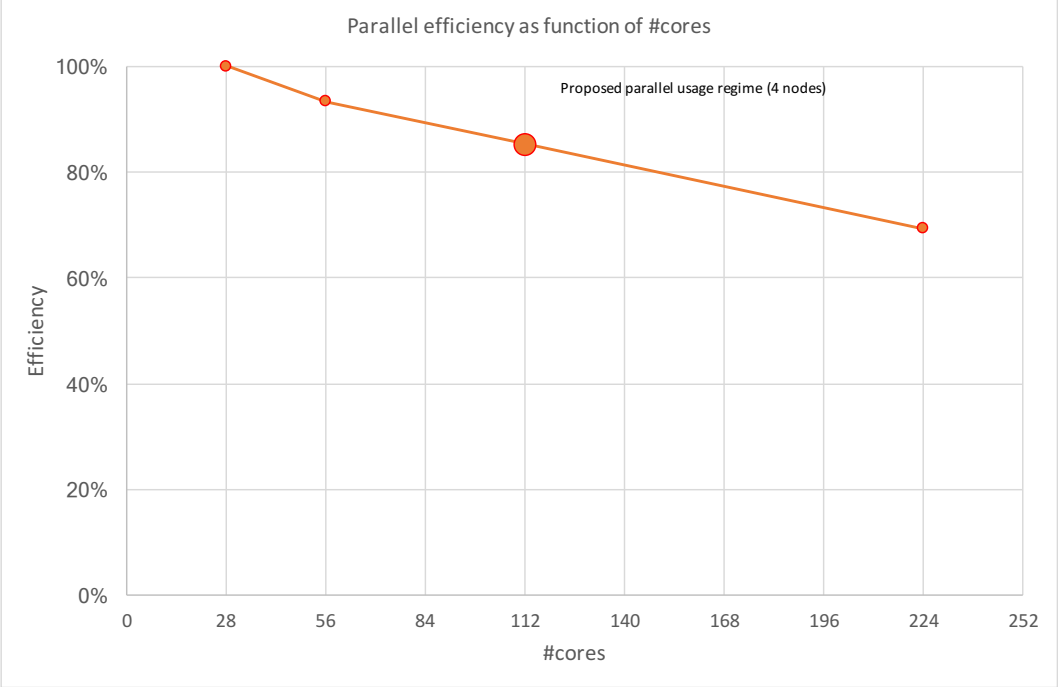
Table 3

# nodes	# cores	Wall clock time (h)	Speedup (with respect to 1 node)	Efficiency
2	28	54.86	1.87	0.933
4	56	31.23	3.28	0.820
8	224	18.14	5.65	0.706

Plot 1


Speedup as function of #cores
Speedup
#cores
0
1
2
3
4
5
6
7
8
9
0
28
56
84
112
140
168
196
224
252
Actual speedup
Ideal speedup

Plot 2


Parallel efficiency as function of #cores
Efficiency
#cores
Proposed parallel usage regime (4 nodes)
100%
80%
60%
40%
20%
0%
0
28
56
84
112
140
168
196
224
252

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, …)? Provide a data management plan (transfer of files to/from Tier1).

The simulations would be primarily performed in the july-august 2017 timeframe, on a regular basis. Job management would be done manually, in order to briefly monitor sanity of the MD simulations in between two subsequent 100-ns runs. In principle, 4 x 8 = 32 simulations could be ran simultaneously, allocating 32 x 4 = 128 nodes, 22% of all BrENIAC computing capacity. If less capacity would be available because of other users, this would impact the throughput time, but given the past Tier-1 usage patterns, we don't predict this delay would be problematic.

Each 100-ns MD run produces 240 Mb of data, which is automatically (i.e. as part of the compute job) transferred from the BrENIAC SCRATCH-space to the DATA partition of Tier-2. This means that, at most, 32 x 240 Mb = 7.68 Gb of Tier-1 SCRATCH-space would be in use. To be on the safe side and never run of out Tier-1 disk quota, we would ask for double that amount: 15.36 Gb.

This entire project would require 76.8 Gb of Tier-2 DATA space, which would be stored there for 3-4 months until analysis has been completed. Given that UGent will bring in production a 3-PB new HOME/DATA storage in the June 2017 timeframe, we don't anticipate this to be an issue. If so, we will regularly offload DATA storage to a local storage volume.

11. List the granted computing time allocations to the promoter(s) of this research project, on the Flemish Tier-1 systems, 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 systems or on other Tier-1 or Tier-0 supercomputers. DOI links are sufficient.

None

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

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: hpc@vub.ac.be
For the other institutions: caroline.volckaert@FWO.be

Beschrijving van een onderzoeksproject

IWETO-PROJECT Inlichtingsformulier - Afdeling Onderzoekscoördinatie

(eveneens beschikbaar via: http://www.UGent.be/nl/onderzoek/activiteiten/iwetoformulier.doc)

Kredietcode: 120C05712W
B/12944/

Promotor: **Tavernier Jan (P3)**

Vakgroep : Biochemie GE07

Adres : **A. Baertsoenkaai 3, 9000 Gent**

Co-promotor: Gevaert Kris, Martens Lennart, Gettemans Jan

Financierende instelling: **Belspo - IUAP VII**

Nederlandse beschrijving

Titel: Analyse van signalisatiecomplexen aan de hand van hedendaagse interactomica

Inhoud (max. 60 woorden):

Interacties tussen eiwitten vervullen een centrale rol in de cel en dysregulatie van eiwit interactienetwerken draagt bij tot diverse humane ziekten. We zullen een unieke technologie “toolbox” die bestaat uit MAPPIT, PPI-iMixPro en VIROTRAP, gecombineerd met state-of-the-art bioinformatica analyse, aanwenden om nieuwe inzichten te verkrijgen in de (patho)fysiologische rol van selecte eiwit interacties. Daarnaast zullen nanobodies als intrabodies gebruikt worden om eiwitinteracties te wijzigen *in cellulo*.

Trefwoorden (vrij te kiezen, min. 3): eiwit interactinetwerken, bioinformatica, nanobodies

Engelse beschrijving

Title: Analysis and mining of signaling complexes using contemporary interactomics

Abstract (max. 60 woorden):

Protein interactions fulfil a critical role in a cell and their dysregulation contributes to human disease. We will apply an unique technology “toolbox” comprising MAPPIT, PPI-iMixPro and VIROTRAP, combined with state-of-the-art bioinformatics approaches, to gain new insights into the (patho)fysiological roles of selected protein interactions. In addition, we will use nanobodies as intrabodies to disturb protein interactions *in cellulo*.

Keywords (vrij te kiezen, min. 3): protein interaction networks, bioinformatics, nanobodies

Disciplinecodes (min. 1)	P310				

Zie: https://www.ugent.be/nl/onderzoek/activiteiten/disciplinecodes.htm

Toepassingscodes (min. 1)	O400				

Zie: https://www.ugent.be/nl/onderzoek/activiteiten/toepassingscodes.htm

Dit formulier moet vóór de start van elk nieuw onderzoeksproject worden ingevuld. IWETO (**I**nventarisatie van het **W**etenschappelijk **e**n **T**echnologisch **O**nderzoek in Vlaanderen) is een gezamenlijk initiatief van de Vlaamse Regering, de Vlaamse Interuniversitaire Raad en de Vlaamse Universiteiten. De IWETO-databank van de

Universiteit Gent wordt beheerd door de afdeling Onderzoekscoördinatie en staat ter beschikking van alle UGent-onderzoekers.

_____UNIVERSITEIT GENT_____
Sint-Pietersnieuwstraat 25, B-9000 Gent Belgium
tel. : ++32-9-264.8921 fax : ++32-9-264.35.83
iweto@UGent.be