eIUS Prototype Use Cases[1]

In these examples, underlined elements are based either on the original interview data or on feedback sent by the informants in response to draft versions of the use cases. Key *activity types* are highlighted in *italics* and **relevant ICTs** in **bold**.

Use Case 3 – Computational Biochemistry

Narrative

1. Martin, a computational biochemist based at Glasgow University is *browsing* the **Protein Data Bank** one morning in December when he comes across a new entry from a colleague based on the East Coast of the USA.

2. Martin knew his X-ray crystallographer friend Thomas was working on determining the cartesian coordinates for this particular protein (human lung cytochrome) but did not realise he was ready to upload the coordinates to the **Protein Data Bank**. He had been waiting for these coordinates because he wanted to run a simulation to *validate some experimental results* recently obtained by another collaborator, Eric, based in the Netherlands.

3. Martin *phones up Thomas* to get the full story, which involves Thomas achieving an unexpected breakthrough in obtaining a viable sample of cytochrome, expected to take many months longer. Thomas did intend to let Martin know earlier but they needed to get the results submitted in a hurry to be included in their research institution's quinquennial review, which has major funding implications.

4. Three years ago, Martin had *performed a similar simulation* of mouse hepatic (liver) cytochrome, using the NAMD molecular dynamics simulation application but no longer had the relevant scripts and input configuration files to hand. Fortunately, Martin was an early depositor to the simulation archive **BioSimGrid**, a fully curated repository of molecular dynamic simulations.

5. Martin *searches* **BioSimGrid** for the keyword, 'cytochrome', and finds that his three-year-old simulation is the second entry in the results. He *clicks on 'download'*, and *saves the* simulation system

---

'mhC-20031003.pdb' along with the accompanying trajectory 'mhC-20031003.xtc' (in the format of the Gromacs, not the original NAMD, molecular dynamics simulation application) to his desktop.

6. By *clicking on this file*, **the molecular viewer, VMD**, is opened immediately. Martin *takes a look* at the protein and the surrounding solvent (water) molecules and *replays the simulation*. The visualisation tallies with what he remembered from the paper he wrote 3 years ago. However, he needs to *replace the mouse hepatic protein with the human lung variant*.

7. Martin then *superimposes the original mouse protein with the human variant*, plugging in Thomas's coordinates from the **Protein Data Bank**, and *submits the new simulation* to his lab's **Local Cluster**.

8. Four hours later, Martin receives a new email with a link to the new simulation set. He *downloads it and replays it*. Unfortunately, the result does not tally with Eric's experimental results.

9. Phoning up Eric, it turns out that the most likely reason for this is because reason for this is because the protonation states for some of the residues in the protein is not simulated realistically in accordance with the physiological pH value. Making the simulation more realistic means further work to the input configuration of the simulation.

10. Eric and Martin work over the next two months to *validate Eric's experimental results* without success. Eventually Eric agrees to *re-do his original experiments*, under more constrained and therefore known conditions.

11. The monitoring of an additional variable through the experiment, leads to the need to alter the simulation to match this, and Martin has to *make a series of modifications to the protein* through the molecular viewer. In particular, the protonation states of some of the residues in the protein and the solvation conditions by water molecules needed to be carefully determined and adjusted, using both existing data in the literature (in the PubMed literature database) and Eric's new results.

12. After two months of work, Martin finally obtains simulation results that tally with experiment and they decide to *publish* a conference paper on the findings.

13. Martin uses **VMD** and its rendering plugins to *create a ray-traced animation* of the results, which is embedded in a presentation he gives to an international biophysics conference in Long Beach. On receiving favourable feedback, Martin and Eric spend the next three months *extending their study* taking into account suggestions.

14. The full results are eventually *written up* within a well-respected computational biochemistry journal.

Relevant ICTs

ICT	Comments
BioSimGrid[2]	A fully curated repository for molecular dynamics simulations (trajectories).
Protein Data Bank[3]	A fully curated repository for the processing and distribution of 3-D structure data of large molecules of proteins and nucleic acids.
Local Cluster	A high performance compute resource provided locally.
VMD	A molecular visualisation program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics. VMD is considered within the scope of eIUS because it has the capability to work with a molecular dynamics program running on another computer in order to display the results of a simulation as they are calculated. However, it is not possible to submit new simulations directly from VMD. New simulations must be submitted manually, demanding a good appreciation of the command line, cluster computing and scripting languages.

---

[2] www.biosimgrid.org
[3] www.pdb.org

Commentary

This use case is an example that illustrates the way in which researchers can interact and combine e-Infrastructure services provided at different levels, from infrastructure provided locally (in the case of local clusters) to infrastructure hosted internationally (in the case of the Protein Data Bank). All the ICTs featured in this use case were given as examples by the two informants based within a computational biochemistry lab at Oxford. Local clusters are highlighted because the informants were both based within a well-funded lab that had three powerful in-house clusters, in addition to a large HPCx[4] allocation. This meant they did not see the benefit of going through the procedures required to gain access to the National Grid Service.

The overall shape of the research described by the use case - comparing experimental results with simulation studies - was taken from the interview data. However, a lack of concrete examples of typical proteins meant an appropriate example was chosen from the computational biochemistry literature.

Comments by Informant

The basic structure and flow of the use case remained unchanged following review. However, the original use case had a series of gaps that were filled in by the informant. These details included:

- in step 4, the incorporation of mouse hepatic (liver) cytochrome as the protein featured in the previous simulation;
- in step 5, the details of the simulation set downloaded from BioSimGrid;
- in step 9, the reason the simulation did not tally with experiment;
- in step 11, the details of how the simulation was altered to match the experiment.

Other Editorial Considerations

Element	Usage
Links to direct quotes?	No
Year?	No
Month?	Yes
Time of day?	Yes
Location given?	No
Real institutions named?	Yes
Real journals named?	No
Real conferences named?	No

---

4 www.hpcx.ac.uk