H2020 EINFRA-5-2015

www.bioexcel.eu

Project Number 675728

D4.5 – Training report and updated plan

WP4: Dissemination and Training

Document Information

Deliverable Number	D4.5
Deliverable Name	Training report and updated plan
Due Date	2017-04-30 (PM18)
Deliverable Lead	EMBL-EBI
Authors	Vera Matser (EMBL-EBI)
Keywords	Training
WP	WP4
Nature	Report
Dissemination Level	Public
Final Version Date	2017-04-28
Reviewed by	Alexandre Bonvin (UU), Adam Carter (UEDIN), Bert de Groot (IMG), Erwin Laure (KTH), Erik Lindahl (KTH), Emiliano Ippoliti (JUELICH), Rossen Apostolov (KTH)
MGT Board Approval	2017-04-29

Document History

Partner	Date	Comments	Version
EMBL-EBI	2017-03-31	First draft, made available for partial feedback	0.1
EMBL-EBI	2017-04-12	Second draft distributed to CoE for feedback	0.2
EMBL-EBI	2017-04-26	Third draft, comments from internal review integrated	0.3
EMBL-EBI	2017-04-28	Fourth draft, comments from PMB integrated and final formatting	0.4

Executive Summary

This document describes the training activities that have been completed during the first period of the BioExcel Centre of Excellence. The BioExcel competency profile underlies the BioExcel training programme and each of our training activities are mapped to the BioExcel competency framework (see Appendix 7.1). The development of the competency framework has been extensively described in previous deliverables D4.2 – Competency framework, mapping to current training & initial training plan and D4.3 – Recommendations for BioExcel Training Programme, and links to zenodo are available on the BioExcel website (http://bioexcel.eu/research/deliverables/).

For the completed face-to-face training events we discuss the event structure and the post-course feedback. To date three BioExcel Training Programme events have taken place and another three are under development. The three completed BioExcel training events have been rated as "Good" or "Excellent" by, respectively, 100%, 95% and 77% of the participants. The BioExcel training programme builds on the already established strong training efforts of each of the BioExcel partners. On top of the efforts of the BioExcel training Programme an additional 28 events with a training element have been reported, reaching a minimum of 766 people in 20 locations.

We present the subset of BioExcel webinars that aim at providing users with the necessary background knowledge needed to take advantage of the BioExcel services; we have named these webinars the educational webinar series. Three educational webinars have been completed, with the total number of BioExcel webinar standing at 13; the webinar programme is coordinated through WP3. More work is needed to promote the online webinar repository and the knowledge resource center (once phase 2 has been launched). This deliverable provides an update on the planning of the upcoming training events and the planned online resources.

The prototype of the second phase of the Knowledge Resource Center is discussed, which has become part of a larger project that includes a competency viewer and will long-term allow users to track their continued professional development based on the competencies.

We provide an update on the Training Interest Group, which currently has 20 members. A kick-off webinar was completed and a number of face-to-face and online activities planned for the interest group over the course of 2017.

Contents

1   Introduction

This document will present the completed BioExcel training activities, discuss plans for upcoming events and update the recommendations for the BioExcel Training programme. It will also report on the status of the generated training material.

Together with D4.4 – Dissemination Report and Updated Plan (PM18), this deliverable gives a complete overview of the work completed under work package 4. D4.4 presents information about the project website, social media channels, dissemination events and publications, while D4.5 focusses on the BioExcel Training Programme. Where content overlaps we will specify if a subset of information is presented and refer to the deliverable with the most detailed overview.

The BioExcel Training programme is built on the BioExcel Competency Framework. The development of the BioExcel competency framework and how it has been used to underpin the training programme has been described in two previous deliverables D4.2 – Competency framework, mapping to current training & initial training plan and D4.3 – Recommendations for BioExcel Training Programme. All BioExcel deliverables are publically available through Zenodo and linked from the BioExcel webpages (http://bioexcel.eu/research/deliverables/).

2   Completed Training Activities

The BioExcel competency profile is the foundation of the BioExcel training programme. Section 2.1 – 2.4 report on the completed BioExcel training events, but it is important to note that these events are not the only training activities coming from the BioExcel partners. The BioExcel training programme builds on the already established strong training efforts of each of the BioExcel partners (section 2.8). The training programme specifically focusses on the training needs areas that have been identified through the gap analysis (see D4.3 – Recommendations for BioExcel Training Programme) where currently available training resources (face-to-face and online) are not providing sufficient coverage.

2.1   Overview
Here we present an overview of the BioExcel training activities that have already taken place (Table 1).

Status	Date	Title
		Competencies (see Appendix 7.1)
Year 1 Completed	3-4th May 2016 (EMBL-EBI)	BioExcel: addressing training needs for advanced simulations in biomolecular research http://bioexcel.eu/events/workshop-addressing-training-needs-for-advanced-simulations-in-biomolecular-research/
		Competency development

Year 1 Completed	20-21st October 2016 (BSC)	BioExcel: workflow training for computational biomolecular research http://bioexcel.eu/events/bioexcel-workflow-training-for-computational-biomolecular-research/
		4.1
Year 2 Completed	10-13th April 2017 (KTH)	BioExcel – PRACE Spring School; Focus on GROMACS and related codes (AMBER, NAMD) and visualization tools (VMD) https://events.prace-ri.eu/event/502/
		2.3, 2.5, 2.6
Year 2 Registration closed	3-7th July 2017 (EMBL-EBI)	Foundation skills for HPC in computational biomolecular research (project based) http://www.ebi.ac.uk/training/events/2017/foundation-skills-hpc-computational-biomolecular-research
		Generic Computing Competencies (3.1 – 3.8)
Year 3	TBC (EPCC)	Introduction to HPC computing for life scientists
		4.2 – 4.5 (building on from "Foundation skills for HPC in computational biomolecular research")
Year 3	TBC	BioExcel Summer School 2018
		2.5; 2.6

Table 1: BioExcel face-to-face training events.

Table 2 provides an overview of the high-level statistics collected for the training events.

Course	Number of attendees	Attendance Rate*	% Feedback	% rated the event Good/Excellent	% would recommend
Event 1	19**	63%	42%	100%	n/a
Event 2	28	93%	79%	95%	91%
Event 3	55	85%	75%	77%	87%

Table 2: Overview of feedback collected for the BioExcel Training events. *number of attendees/number of available spaces, **includes speakers due to knowledge exchange nature of the event.

Each of the completed face-to-face events will be discussed in detail. For the events which have not been completed an update is provided in section 3. In addition we will discuss the travel grants, the BioExcel educational webinar series and provide an overview of the additional training activities that the BioExcel partners have been involved in.

2.2 BioExcel: addressing training needs for advanced simulations in biomolecular research

The first BioExcel Training event "BioExcel: addressing training needs for advanced simulations in biomolecular research" was held on 3rd -4th May 2016 at the European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.

2.2.1   Course Overview

The purpose of this workshop was twofold; firstly, it provided users of three important biomolecular modelling and simulation resources, GROMACS, HADDOCK and CPMD, to learn directly from the developers of these resources and find out about how BioExcel - Europe's new centre of excellence for BioMolecular Research - will enable users to harness high-performance computing (HPC) to use these tools in a user-friendly environment that does not assume any prior experience of using HPC. 'Surgeries' enabled participants to have one-to-one discussions with the developers of these resources, so that together they could discuss bottlenecks in their modelling and simulation workflows.

Secondly, it enabled users of these services to provide BioExcel with feedback on the competencies needed to fully exploit these services. The methods employed and the outcomes of this second part of the training event has been extensively described in previous deliverables D4.2 – Competency framework, mapping to current training & initial training plan and D4.3 – Recommendations for BioExcel Training Programme. This report will focus on the end-user training aspect and the course feedback of this event.

2.2.2   Audience

The workshop was aimed at existing users – both academic and industrial – of GROMACS, HADDOCK and CPMD. Participants were made aware in advance that, though end-user training was included in this workshop, it focused on knowledge exchange and rather than being a classical training workshop. The aim was to attract participants who had expertise in molecular modelling in general and in at least one of the flagship software applications in BioExcel, in addition to some proficiency in Linux. We advertised the workshop through the EMBL-EBI Industry Programme to attracting participants from both academia as well as industry. Participants were expected to contribute to the discussion around use cases and training requirements for BioExcel.

2.2.3   Outcomes

From participant's point of view:

- Learn about what GROMACS, HADDOCK and CPMD can currently do for your research
- Find out about what BioExcel is planning and how the centre will support your research in the future
- Get to know core developers and provide input into the future development of the applications
- Share where your bottlenecks are so that we can develop appropriate, more automated workflows for accelerating your research
- Provide input into your future training needs in this topic, so that we can deliver relevant and high quality training to our users

From BioExcel's point of view:

- Input and feedback on BioExcel user groups and use cases
- Gather input on the competencies needed to fully exploit BioExcel services
- Promote BioExcel CoE and specifically the training programme

2.2.4 Programme

Time	Topic	Trainer
Day 1 – Tuesday 3rd May 2016		
09:00 – 09:30	Arrival and registration	
09:30 – 09:45	Welcome and warm up activity	Cath Brooksbank and Vera Matser, EMBL-EBI
09:45 – 10:15	What is BioExcel and what can it do for you?	Rossen Apostolov, KTH Royal Institute of Technology
10:15 – 10:35	Application overview 1: Haddock	Alexandre Bonvin, Utrecht University
10:35 – 10:55	Application overview 2: GROMACS	Mark Abraham, KTH Royal Institute of Technology
10:55 – 11:25	**Break**	
11:25 – 11:45	Application overview 3: CPMD	Emiliano Ippoliti, Forschungszentrum Jülich
11:45 – 12:45	Interactive exercise - defining user groups and use cases	Vera Matser, EMBL-EBI and Adam Carter, EPCC
12:45 – 13:45	**Lunch**	
13:45 – 15:30	Hands-on tutorials and surgery sessions	GROMACS, HADDOCK and CPMD developers
15:30 – 16:00	**Break**	
16:00 – 17:45	Hands-on tutorials and surgery sessions continue	
17:45 – 18:00	Wrap up of day 1	Cath Brooksbank
18:00	**Drinks and dinner at Hinxton Hall restaurant**	
Day 2 – Wednesday 4th May 2016		
09:00 – 09:30	Intro to BioExcel's competency-based training concept and draft competency profile	Cath Brooksbank and Vera Matser
09:30 – 10:30	Use cases: three talks from users, illustrating their research challenges	Speakers to be selected from registrants
10:30 – 11:00	**Break**	
11:00 – 12:00	Breakout 1: what competences are needed; what's not needed and what's provided by other training programmes?	ALL
12:00 – 12:30	Reports from breakouts	Breakout rapporteurs

12:30 – 13:30	**Lunch**	
13:30 – 15:30	Hackathon: compiling input from the breakouts and adding knowledge, skills and behaviours for each competency	ALL
15:30 – 16:00	**Break**	
16:00 – 16:20	Presentation of outcomes	Cath Brooksbank and Vera Matser
16:20 – 17:30	Final use-case surgery	GROMACS, HADDOCK and CPMD developers
17:30	Wrap up and departures	

Table 3: Programme for "BioExcel: addressing training needs for advanced simulations in biomolecular research", 3-4 May 2016, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.

2.2.5   Conclusion and course feedback

A total of 19 people attended the event, this includes delegates and speakers, though this distinction is somewhat artificial in this event. The attendance rate for this event was 63% (places filled / maximum places available).

A combination of reasons are responsible for this low percentage; it is notoriously difficult to attract delegates to an event where they will not necessarily gain skills or knowledge but are mainly being asked to contribute. When this event was organised, the Centre of Excellence was still relatively unknown and did not yet have well established dissemination strategy to promote event, therefore it was more difficult to attract delegates to this event. Under ideal circumstances the event would have been opened for registration substantially earlier. However, the group of delegates that was gathered were very engaged and provided us with high quality feedback on the BioExcel user groups, use cases & competency framework.

Note 1: Low attendance at events has been added to the BioExcel Risk register as one of the major risks of WP4 (R16 – risk score 15, category High). Low attendance is a risk identified by many training organisations, not just BioExcel. In part this may be related to the financial burden of attending a training course, even if the training course does not charge a registration fee. BioExcel has decided to offer a number of fixed-amount travel grants for each training event to allow participants to attend regardless of financial circumstances. The travel grants will be awarded on a competitive basis.

The following mitigation actions have been added to the risk management plan: Effective promotion of events beginning several months before the event. Ideally registration should be open 6 months in advance with promotion of the event starting earlier or at the same time. Developing training that is specifically aimed towards filling perceived (and real) gaps in training. Travel grants to mitigate financial burden of attending training events.

"Number of people who have attended BioExcel courses" has been identified as one of the WP4 KPIs (WP4_KPI03), the reachable target has been set at 80% and the stretch target at 100%.

Based on the evaluation of the risks and KPIs at PM18 we can conclude that our mitigation actions are working. Table 2 shows the attendance rate for the completed events, the next BioExcel event is expected to be fully booked (see section 3.2.1). We will continue to expand our dissemination network to help increase the reach of the CoE.

A standardised post-course feedback survey is used to collect course feedback for all BioExcel training event. Feedback percentage for this course was not very high 42% (8/19), though within acceptable limits (see note 2). Looking at the respondents this is mostly the external delegated, meaning non-BioExcel and non-EMBL-EBI (7/19), which is good. The course feedback was very positive, 100% of the respondents rated the overall workshop as Good or Excellent and indicated that their expectations were either "Met" or "Exceeded". The hands-on tutorials & meeting the developers of the codes were seen as the most useful part of the workshop (6/8). Two delegates indicated that they didn't find the creation of the user stories or respectively the discussion of the competencies very useful to them (but they enjoyed the social aspect of it). Relatively this is not surprising since it is of direct benefits to the BioExcel CoE but only of indirect benefit to the delegates. When asked whether they would be interested to receive further training on this topic 7 out of 8 respondents answered "Yes"; the highest requested topic was advanced or additional training in one or more of the BioExcel codes.

Note 2: WP4_KPI04 looks at course feedback, it specifies a reachable target of 60% responses in the Good/Excellent category with 80% as the stretch target. Both are based on an expected response rate of 30-50%. Asking people to fill in the feedback form during the event will help to increase the response rate.

2.3 BioExcel: Workflow Training for Computational Biomolecular Research

The second BioExcel Training event "BioExcel: Workflow Training for Computational Biomolecular Research" was held on 20th – 21st October 2016 at the Barcelona Supercomputing Center in Spain. This event was jointly organised with the WP3 Workflows Interest Group.

2.3.1 Course Overview

The purpose of this workshop was to:

- Provide an overview of the most commonly used workflow managers in the computational biomolecular field, both in a theoretical and practical way, with hands-on practice sessions.

- For participants to find out about how BioExcel - Europe's new Centre of Excellence for Computational BioMolecular Research - will help them to develop and run their own pipelines using these tools and introduce them to the BioExcel Workflows Interest Group

- A "Bring Your Own Workflow" session allows participants to identify the best approach for their particular workflow problem with the help of experts on the different workflow managers.

2.3.2 Audience

Prior knowledge of workflow managers was not required for this event but participants were strongly encouraged to make sure that they have a sufficiently clear idea of the bioinformatics problem they were trying to solve and be able to express this as a simple workflow on paper. Participants were encouraged to bring their own data or problem statement to the workshop. No software installation was necessary, a Virtual Machine (VM) was provided on the day.

2.3.3 Learning outcomes

The following learning outcomes were defined for the workshop:

After the course participants should be able to:

- Explain what Workflow Managers are, in which way they differ from each other, and how you can use them in your research.

- Anticipate how BioExcel might be able to support participants in his/her future research and know what BioExcel is planning with regards to Workflow Managers.

- Make an action plan to implement or refine his/her own workflow, based on one-to-one discussions with the Workflow Managers experts.

2.3.4   Programme

Time	Topic	Trainer
Day 1 – Thursday 20th October 2016		
08:30 – 09:00	Arrival and registration	
09:00 – 09:30	Welcome and expectations post up	Vera Matser, EMBL-EBI
09:30 – 10:00	What is BioExcel and what can it do for you?	Rossen Apostolov, KTH Royal Institute of Technology
10:00 – 10:30	State of the art of Workflow management?	Stian Soiland-Reyes, UNIMAN
10:30 – 11:00	**Break**	
11:00 – 12:30	Galaxy + Hands-on	Antonio Miguel Espinosa Morales, UAB Barcelona
12:30 – 14:00	KNIME + OpenPHACTS & Hands-on	Daniela Digles, Universität Wien
14:00 – 15:00	**Lunch**	
15:00 – 15:45	Copernicus + Hands-on	Mark Abraham, KTH
15:45 – 16:30	Nextflow	Paolo Di Tommaso, CRG
16:30 – 17:30	Taverna + Hands-on	Stian Soiland-Reyes, UNIMAN
17:30 – 18:00	Marenostrum Visit	
19:00 – 21:00	**Workshop Dinner**	
Day 2 – Friday 21th October 2016		
09:00 – 09:45	Common Workflow Language (CWL)	Michael Crusoe, Brussels, Belgium
09:45 – 10:30	BioExcel Workflows Interest Group	Darren White, EPCC
10:30 – 11:00	**Break**	
11:00 – 11:30	BcBIO	Darren White, EPCC
11:30 – 13:00	PyCOMPSs & Hands-on	Daniele Lezzi, BSC / Pau Andrio, BSC
13:00 – 14:00	**Lunch**	
14:00 – 18:00	Bring Your Own Workflow!	All
18:00	End of course	

Table 4: Programme for "BioExcel: Workflow Training for Computational Biomolecular Research", 20–21 October 2016, Barcelona Supercomputing Center (BSC), Spain.

2.3.5   Conclusion and course feedback

A total of 28 people attended the event, this translates to a 93% attendance rate (28/30). Due to a last minute cancellation it was not possible to offer the place to anyone else. We received course feedback from 22 delegates, which is a response

rate of 79% (22/28). The event was very successful, well attended (see figure 1) and a range of different type of participants attended. The largest group was PhD students, just over 30%. Unusually 30% of delegates indicated the "other" option on "Type of delegate"; looking into detail this includes industry representatives, and software or computer engineers.

95% of the respondents rated the overall workshop as "Good" or "Excellent". Responses for the best part of the workshop cover most of the main elements of the intended aim of the workshop; hands-on sessions of the different workflow managers, discussion with peer and experts, presentation of the different systems/current state-of-the-art.

The main aspect that was flagged as having potential for improvement was more time for the hands-on sessions, mainly to counteract the fast pace of the workshop. When asked about the balance between theoretically and practical content this aspect came across as well, with just under 60% of the respondents saying the balance was too far towards the theoretical (40% said "About right"). Based on this we can explore the options of another course with more time for hands-on or to expand the online offering for workflows, either through BioExcel or by pointing users at external resources. 55% of respondents preferred online training when asked about which training offerings they would prefer to attend. The majority of respondents (~60%) had occasionally used a subset of the workflow systems before. What is a little bit worrying is that just over 80% of the respondents replied "Maybe" to the question "Will you use these resources now?", the other 20% said "Yes". Over 90% of the respondents said they would recommend future BioExcel training events to their colleagues. Combining this information with the overall rating of the workshop, it is clear that the respondents found the event useful. It is hard to tell exactly why this does not translate to a higher number of respondents planning to use the resources in future. We may be able to get a clearer idea of this in the long-term feedback survey (see section 3.5).

Workflow Training for Computational Biomolecular Research
20 - 21st October 2016 - BSC, Barcelona, Spain

Figure 1: course photo as sent to the participants after the event.

2.4 BioExcel/PRACE Spring School 2017 – HPC for Life Sciences

The third BioExcel Training event, jointly organised with PRACE, "BioExcel/PRACE Spring School 2017 – HPC for Life Sciences" is taking place on 10th – 13th April 2017 at KTH in Stockholm, Sweden.

2.4.1 Course Overview

For the third training event "BioExcel/PRACE Spring School 2017 – HPC for Life Sciences" BioExcel Centre of Excellence has brought together, for the first time, the main developers of four of the major and most popular codes for molecular modelling and simulations - GROMACS, AMBER, NAMD and VMD. It has given the participants, whether they are novice or experienced users a great opportunity to learn from the very authors of the software.

The 2017 Spring School was organised jointly with PRACE, the main HPC resource provider in Europe. The school program allowed participants to get a comprehensive introduction to the different codes; to understand their scalability and performance, as well as ways to avoid potential issues; to learn the best practices about using them on HPC systems. Extensive hands-on sessions covered more than half the time of the event.

The course took place at the computer laboratories of PDC Center for High Performance Computing at KTH Royal Institute of Technology in Stockholm. Participants were able to run on the PRACE Tier-1 system Beskow, the largest Tier-1 system within the Nordic countries.

The school was also supported by the Swedish National Infrastructure for Computing (SNIC).

2.4.2 Programme

Time	Topic	Trainer
Day 1 – Monday 10th April 2017		
08:30 – 09:00	Arrival and registration	
09:00 – 09:15	Welcome	Michaela Barth, KTH
09:15 – 10:00	BioExcel Center of Excellence - Training Programme	Vera Matser, EMBL-EBI
10:00 – 10:30	Break	
10:30 – 12:00	Molecular visualisations with VMD	John Stone
12:00 – 13:00	Lunch	
13:00 – 13:45	How to use the PDC Supercomputing systems	Henric Zazzi
13:45 – 15:00	Hands-on VMD	
15:00 – 15:30	Break	

15:30 – 17:30	Hands-on VMD	
Day 2 – Tuesday 11ᵗʰ April 2017		
09:30 – 10:30	HPC Simulations with NAMD	Jim Phillips
10:30 – 11:00	Break	
11:00 – 12:00	HPC Simulations with NAMD	
12:00 – 13:00	Lunch	
13:00 – 14:30	Hands-on NAMD	
14:30 – 15:00	Break	
15:00 – 17:00	Hands-on NAMD	
Day 3 – Wednesday 12ᵗʰ April 2017		
09:30 – 10:30	HPC Simulations with GROMACS	Erik Lindahl, KTH
10:30 – 11:00	Break	
11:00 – 12:00	HPC Simulations with GROMACS	
12:00 – 13:00	Lunch	
13:00 – 14:30	Hands-on GROMACS	
14:30 – 15:00	Break	
15:00 – 17:00	Hands-on GROMACS	
18:00	Social Event Dinner	Vasa Museum
Day 4 – Thursday 13ᵗʰ April 2017		
09:30 – 10:30	HPC Simulations with AMBER	Tom Cheatham
10:30 – 11:00	Break	
11:00 – 12:00	HPC Simulations with AMBER	
12:00 – 13:00	Lunch	
13:00 – 14:30	Hands-on AMBER	
14:30 – 15:00	Break	
15:00 – 17:00	Hands-on AMBER	

Table 5: Programme for "BioExcel/PRACE Spring School 2017 – HPC for Life Sciences", 10–13 April 2017, KTH, Stockholm, Sweden.

2.4.3   Conclusion and course feedback

The Spring School was attended by 55 people, giving us an attendance rate of 85%. It should be noted that the school took place a few days after the terror attack in Stockholm and a number of people therefore cancelled their participation at short notice. 75% of the participants filled out the feedback survey, though a reminder has been sent to try to increase that number. 77% of the respondents rated the event as "Good" or "Excellent", this number is in line with the reachable target defined by WP4 but is sufficiently low that specific attention needs to be given to the post-event evaluation with the organisers to identify why the percentage is lower that we have seen in previous courses.

The main reasons for lower numbers had already become clear during the event itself. The first reason is that there were substantial problems with the login/creation of the user account. Though problematic for the hands-on component of this event, the problem appears to be the result of a change in procedures in setting up the user account which did not filter down to the event organisers and should therefore not occur again. A separate reason listed by some of the respondents is that the organisation of the hands-on sessions (separate

from the login problems) were disorganised and that insufficient help was available. To which degree this problem was felt varies across the different hands-on session.

The main reason for the lower rating, which is also in part responsible for the problems with the hands-on session, is that the audience was the Spring School was ill defined. The level of expertise ranged from beginners with no expertise in any of the four codes to participants who were proficient in one or more of the codes. In hindsight the range of proficiency was too great with result that the participants without prior knowledge felt lost and not sufficiently supported. To some extent this problem was made worse by the fact that their level of expertise was asked for but subsequently not acted upon sufficiently. Similar to the hands-on support problem the degree to which this was felt to be a problem varies between the sessions. The speakers of day 3 & day 4 were made aware of the frustration of the beginner participants which allowed them to take this into account.

We are planning to liaise with some of the beginner participants and the Entry Level Users Interest Group to provide better support to this user group, specifically related to the BioExcel codes.

Respondents viewed the best part of the course to be the opportunity to have the codes presented side by side by their respective lead developers. The lectures by Erik Lindahl (GROMACS) and John Stone (VMD) were also considered highlights. For a BioExcel point of view, bringing together, for the first time, the main developers of four of the major and most popular codes for molecular modelling and simulations was a major achievement.

A full debrief with the organisers is planned for early May.

2.5   BioExcel Travel Grant

BioExcel offers a limited number of fixed amount travel grant for each of the BioExcel Training events. Full terms and conditions are made available to prospective applicants on the events page of each of the BioExcel Training courses. When selecting the travel grant beneficiaries we generally take into account gender, geographic spread, and how much impact the course would have on the applicant's research.

Table 6 provides an overview of the travel grants that have been awarded up to this point in time.

Last Name	First Name	Company Name
20-21st October 2016 - BioExcel: workflow training for computational biomolecular research		
Camacho Cano	Esther	Centro de Biología Molecular Severo Ochoa
Flores	Antonio	Universidad de Málaga

Hernandes Coutinho*	Felipe *	Universidade Federal do Rio de Janeiro *
Matsoukas	Minos	University of Patras
Gurinova	Jana	University of Vienna
10-13th April 2016 - PRACE/BioExcel Spring School - HPC for Life Science		
Kravchuk	Vladyslav	Taras Shevchenko National University
Llorach Pares	Laura	University of Barcelona - Mind the byte
Rutkowska	Ewelina	University of Warsaw

Table 6: List of recipients of a BioExcel Travel Grant. *Due to unforeseen circumstances this person needed to cancel their participation in the training event and subsequently the grant was withdrawn.

The beneficiaries of the BioExcel travel grant are asked to write a blog post for us, either about their experiences at the course, their research interests (in layman's terms) or a relevant scientific topic. Appendix 7.3 has the blogs from the second training event "BioExcel: workflow training for computational biomolecular research", which were publicised on the BioExcel website, and shared in the BioExcel newsletter and on twitter.

2.6 Training webinar series

The BioExcel Dissemination and Training work package contributes webinars to the main BioExcel webinar programme, which is organised through WP3. We market these training webinars as the BioExcel educational series to emphasize that these webinars have a different target audience compared to the standard BioExcel webinars and focus on the fundamental background knowledge needed to take advantage of the services provided by BioExcel.

Table 7 provides an overview of the educational webinars that have taken place to date. The total number of webinars organised by WP3 is 13, D4.4 provides a full overview of the webinar series.

Date	Title or description & Speaker		
Competency	**Registrants**	**Attendees**	**YouTube views**
08/02/2017	Assessing structure quality in the PDB archive (Speaker: Matthew Conroy, EMBL-EBI)		
2.3, 2.5	24	11	51
08/03/2017	How to choose compute resources for your team (Speaker: Lee Larcombe, ELIXIR-UK)		
3.2	21	11	29
15/03/2017	Accessing Cloud & HTC Resources in BioExcel (Speaker: Steven Newhouse, EMBL-EBI).		
2.7, 3.2, 4.4	13	3	n/a*

Table 7: Overview of the BioExcel educational webinar series. *Note: there was a technical problem with recording of this webinar; the video will be made available after a new recording has been made.

The number of registrants attending the educational webinar series has been lower than we would like. We do also see this in the main BioExcel webinar series. A separate action plan will be developed to more actively promote the online resources offered by BioExcel.

2.7 Training material

The following training material was generated during the first reporting period of the BioExcel Centre of Excellence.

2.7.1 University of Utrecht:

All tutorials from the Bonvin lab are available through the education page of the laboratory web pages as well as through the BioExcel web page / Knowledge Resource Centre.

Tutorials and training material

HADDOCK basic protein-protein docking tutorial: A tutorial demonstrating the use of the HADDOCK web server to model a protein-protein complex using interface information derived from NMR chemical shift perturbation data. This tutorial does not require any Linux expertise and only makes use of our web server and PyMol for visualisation/analysis.

HADDOCK ab-initio, multi-body symmetrical docking tutorial: A tutorial demonstrating multi-body docking with HADDOCK using its ab-initio mode with symmetry restraints. It is based on a former CASP-CAPRI target (T70).

HADDOCK ligand binding site tutorial: A tutorial demonstrating the use of HADDOCK in ab-initio mode to screen for potential ligand binding sites. The information from the ab-initio run is then used to setup a binding pocket-targeted protein-ligand docking run. We use as example the multidrug exporter AcrB.

PowerFit tutorial: A small introduction into PowerFit to automatically place a high-resolution atomic structure in a lower-resolution cryo-electron microscopy density map.

PowerFit web server tutorial: A small introduction into PowerFit to automatically place a high-resolution atomic structure in a lower-resolution cryo-electron microscopy density map making use of our web portal (does not require Linux).

DisVis web server tutorial: A small introduction into DisVis to analyse the interaction space between two molecules from a set of restraints. It can help to filter out putative false positive restraints and predict key residues involved in the interaction from this set of restraints. It makes use of our web portal (does not require Linux).

[Tutorial homology modelling / molecular dynamics using Gromacs and docking with HADDOCK](): Tutorial created as part of the Structural Bioinformatics & Modelling course of the Master programme of University of Utrecht.

Training related publications

J.P.G.L.M. Rodrigues, A.S.J. Melquiond and A.M.J.J. Bonvin. [Molecular Dynamics Characterization of the Conformational Landscape of Small Peptides: A series of hands-on collaborative practical sessions for undergraduate students](). *Biochemistry and Molecular Biology Education*, *44*, 160-167 (2016).

C. Geng, S. Narasimhan, J. P.G.L.M. Rodrigues and A.M.J.J. Bonvin. [Information-driven, ensemble flexible peptide docking using HADDOCK](). *Methods in Molecular Biology: Modeling Peptide-Protein Interactions.* Eds Ora Schueler-Furman and Nir London. Humana Press Inc. *1561*, 109-138 (2017).

2.7.2   Max Planck Institute

Tutorials and training material
[pmx webserver:]() newly launced pmx webserver aimed at the novice user of alchemical free energy calculations
[pmx tutorial](): step-by step example of how to carry out mutation free energy calculations using pmx and gromacs
[pmx webserver tutorial:]() quick instructions on how to navigate the pmx webserver

Training related publications

The pmx webserver is described in the following recent publication:

V. Gapsys, and B.L. de Groot [pmx Webserver: A User Friendly Interface for Alchemistry]() J. Chem. Inf. Model., 57 (2), pp 109–114 (2017).

2.8   BioExcel Partner training activities

Above we have described the training activities that are coordinated through the BioExcel Dissemination and Training work package, however, most BioExcel partners organise or participate in other training activities through their group, institutes or through other relevant projects where they present work directly funded or of relevance to BioExcel. An additional 28 events with a training element have been carried out by the BioExcel partners. These events have taken place in 20 locations across Europe and Asia and have reach a minimum of 766 people. Since we currently do not have number for each event we know they actual number of people reached will be much higher.

Table 8 presents an overview of the training activities that have taken place to date.

Date	Event name	Location	# Participants (if available)
4 February 2016	PATC Course: Programming Distributed Computing Platforms with COMPSs	Barcelona, Spain	25
29 February 2016	Workshop: In silico tools in drug design and target discovery	Barcelona, Spain	120
14-15 March 2016	PRACE Course. Simulation Environments in Life Sciences	Barcelona, Spain	26
5 - 22 April 2016	HPC-LEAP School on numerical analysis and algorithms at the exascale: Classical N-body methods for complex systems on massively parallel architectures: CoS-1	RWTH Aachen, Germany & FZJ Juelich, Germany	40
11-15 April 2016	INSTRUCT practical course on "Advanced methods for the integration of diverse structural data with NMR data"	Utrecht, Netherlands	20 PhD students
3 May 2016	EMBL-EBI workshop on addressing training needs	Hinxton, UK	15
12 May 2016	ETP4HPC workshop	Prague, Czech Republic	100
16-17 May 2016	ExTASY Tutorial: Tools for Advanced Sampling of Macromolecular Systems, 16/17 May 2016	Edinburgh, UK	35
16-20 May 2016	PRACE Spring School and E-CAM tutorial on Molecular & Atomic Modelling	University College Dublin, Ireland	35
19-20 May 2016	Workshop: GROMACS on highly parallel and heterogenous systems	Gottingen, Germany	50
27-28 May 2016	Workshop: Computer Simulation and Theory of Macromolecules	Hunfeld, Germany	40
6-10 June 2016	School on Molecular Modelling for Life Sciences	Pula, Sardinia, Italy	
15-17 June 2016	BITS 2016: 13th Annual Meeting of the Bioinformatics Italian Society	Salerno, Italy	25 in workshop, 140 overall
27-29 June 2016	6th International Conference on the Development of Biomedical Engineering in Vietnam	Ho Chi Minh City, Vietnam	50
30 June - 1 July 2016	Workshop: Simulating molecular dynamics for drug design	Hanoi, Vietnam	30
4-9 July 2016	EMBO course on integrative modelling	Barcelona, Spain	
20-21 July 2016	Institute Pasteur course on integrative structural biology	Paris, France	40
1-3 August, 2016	Hybrid approaches to describe molecular machines: Introduction to molecular modelling for experimentalists	Vienna, Austria	
3-7 September 2016	ECCB	The Hague, Netherlands	

17-19 October 2016	GROMACS Hackathon 2016	Barcelona, Spain	
19-21 October 2016	VI-SEEM Life Sciences regional training	Belgrade, Serbia	
20-21 October 2016	BioExcel: workflow training for computational biomolecular research	Barcelona, Spain	
24 October - 4 November 2016	HPC-LEAP Workshop: Computational biology and applications in human health	Nicosia, Cyprus	20
7-11 November 2016	Helmholtz Training Course on Integrative Structural Biology	Braunschweig, Germany	25
2 February 2017	Programming distributed computing platforms with COMPSs @BSC	Barcelona, Spain	25
5-10 March 2017	IGSC 2017	Taipei, Taiwan	
14-15 March 2017	PRACE Course. Simulation Environments in Life Sciences	Barcelona, Spain	35
24-25 March 2017	Hunfeld Workshop 2017	Hunfeld, Germany	
3-7 April 2017	Pasteur Course on Bioinformatics	Paris, France	
10-13 April 2017	BioExcel/PRACE Spring School 2017 – HPC for Life Sciences	Stockholm, Sweden	65
14 April 2017	Satellite Meeting on "Recent Advances in computational and experimental molecular biophysics	Ho Chi Minh City, Vietnam	25

Table 8: List of events with a training component organised or with contributions of BioExcel partners.

3   Updated Training Plan

Here we present an overview of the planning for the remaining face-to-face events, updated plans for the online activities and guidelines for BioExcel events. The Knowledge Resource Center and Training Interest Group are discussed separately in sections 4 and 5.1.

3.1   Foundation skills for HPC in computational biomolecular research - BioExcel Summer School

The fourth BioExcel training event "Foundation skills for HPC in computational biomolecular research - BioExcel Summer School" will take place on 3-7th July 2017 at the European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.

3.1.1   Course overview
The 5-day BioExcel Summer School will use project-based learning to empower life scientists to get the most out of using computers. Using example problems and challenges that life-scientists will come across in their day-to-day work, e.g. obtaining data, installing software, cleaning data, analysing data; this course will

help participants gain the foundational computational skills needed to move into the domain of high-end computing.

3.1.2 Audience

The course is a primer for people who want to start using High Performance Computing (HPC) but find the concept quite daunting and lack the needed computational skills. This summer school is project-based, formal classroom teaching will be kept to a minimum. No prerequisites are required for this course. Participants from any career stage are welcome to apply for this course.

Based on the current level of registration received we expect this event to be fully book. We will operate a waiting list for applicant who were not successful in the initial selection. The maximum number of applicants for this event is 25.

3.1.3 Syllabus, tools and resources

The summer school will explore concepts such as computer architecture, hardware and software requirements, operating system, file structure, non-trivial software installation, using Linux commands, automating tasks using the Unix shell, software documentation & best practice.

3.1.4 Outcomes

After this course delegates will be able to or have acquired:
- Confidently navigate through the file structure and operating system of the computer.
- Working knowledge of Linux, including basic command line operations
- Obtain & submit data from/to a range of external sources
- Compile, install and test software
- Read and modify scripts

What will this course not do?

- This course will not teach delegates how to programme, though they would be much better prepared after this course to start a self-taught course
- This is not a course on HPC computing but will teach delegates the foundational skills and concepts that they need to start learning about HPC computing. A full course "Hands-on Introduction to HPC for life-scientists" will be organised in autumn 2017 or early 2018.

3.1.5 Programme

Time	Topic	Trainer
Day 1 – Monday 3rd July 2017		
10:00 – 11:00	Introduction to EMBL_EBI and BioExcel	Vera Matser, EMBL-EBI

11:00 – 11:30	Course expectations	Vera Matser, EMBL-EBI
11:30 – 12:00	User stories	
12:00 – 13:00	Lunch	
13:00 – 15:00	The computational roadmap; from your local machine to clusters and HPC	
15:00 – 15:30	Break	
15:30 – 17:30	Introduction to project work	
18:00 – 20:00	Evening meal (exact time and venue tbc)	
Day 2 – Tuesday 4th July 2017		
09:00 – 09:30	Summary of Day 1 and introduction to concepts of Day 2 using the computational route map	
09:30 – 12:30	Project 1	
12:30 – 13:30	Lunch	
13:30 – 16:30	Project 2	
16:30 – 17:30	Keynote 1	
18:00 – 20:00	Evening meal (exact time and venue tbc)	
Day 3 – Wednesday 5th July 2017		
09:00 – 09:30	Summary of Day 2 and introduction to concepts of Day 3 using the computational route map	
09:30 – 12:30	Project 3	
12:30 – 13:30	Lunch	
13:30 – 16:30	Project 4	
16:30 – 17:30	Visit to EMBL-EBI data centre	
18:00 – 20:00	Evening meal (exact time and venue tbc)	
Day 4 – Thursday 6th July 2017		
09:00 – 09:30	Summary of Day 3 and introduction to concepts of Day 4 using the computational route map	
09:30 – 12:30	Project 5	
12:30 – 13:30	Lunch	
13:30 – 17:30	Project 6	
18:00 – 20:00	Evening meal (exact time and venue tbc)	
Day 5 – Friday 7th July 2017		
09:00 – 09:30	Summary of Day 4 using the computational route map	
09:30 – 13:00	Introduction to HPC computing	
13:00 – 14:00	Course feedback and wrap up	

14:00	End of course (lunch vouchers will be provided)

Table 9: Programme for "Foundation skills for HPC in computational biomolecular research - BioExcel Summer School", 3-7 July 2017, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.

3.2   Hands-on Introduction to HPC for Life Scientists

The working title for the fifth BioExcel training event is "Hands-on Introduction to HPC for Life Scientists". It will likely take place in November 2017 in Edinburgh and will be jointly organised with PRACE. The course will be 3-4 days long and can be seen as a follow up to Foundation skills for HPC in computational biomolecular research - BioExcel Summer School (3-7th July, 2017 – Hinxton, UK). We are currently in the process of confirming dates based on availability of training rooms, though we are considering the option of delegates bringing their own laptops.

It will not be necessary for delegates to have attended the BioExcel Summer School but the learning outcomes attached to the course will be the prerequisites to the autumn hands-on introduction to HPC. If potential delegates have those skills they can directly attend the Hands-on to HPC course. The course is aimed at a level where delegates who have attended the Summer School and have continued to develop their skills will be able comfortable to follow along.

The hands-on introduction course will teach similar material to previous run courses at EPCC but rewritten to a slower pace and with examples and use cases relevant to life scientists.

The vision is that life scientists, with limited prior computational experience, who will attend the foundational skills and hands-on introduction course should be ready to start using HPC facilities and attend further courses by other providers. Currently many of the other courses provided are not specifically aimed at life scientists or require prerequisites (such as programming) and/or advance at a speed that is prohibitive. We will investigate the option of making part or all of the material created for these two events available online to allow on-demand access.

We aim to open registration for this event in May 2017.

3.3   BioExcel Summer School 2018

The sixth BioExcel Training event will be another summer school. Plans for the BioExcel Summer School 2018 are in the early stages, the likely format will be a 4-5 day summer school integrating the full spectrum of BioExcel activities tied together by a use case. Drug design is being discussed as a potential use case. We are in discussions with Sardegna Ricerche, managers of the Science and Technology Park of Sardinia, located in Cagliari, Sardinia in Italy.

A preliminary programme could be lectures in the morning and hands-on in the afternoon covering the following topics:

- Introduction (covering BioExcel, use case etc)
- Docking / HADDOCK
- MD general aspects
- Enhanced sampling techniques (e.g. meta dynamics)
- Free energy calculations
- PMX / protein-ligand related aspects

Additionally the summer school could include the following activities
- Poster session(s)
- Flash talks by participants
- Social event

3.4 Online & blended learning

The following online and blended-learning activities are under discussion for the second reporting period (table 10). Implementation of the activities listed in this table will be dependent on the capacity and availability of the BioExcel partners. It is unlikely that all of these activities will be completed during the current funding period of the CoE.

Title or description	
Type	**Competency**
Introduction webinar to Common Workflow Language (CWL) (Speaker: Michael Cruseo, CWL Project co-founder & Community Engineer)	
Webinar - 31st May 2017	4.1
An entry level webinar(s) that introduces users to concepts underlying biomolecular simulations (e.g. docking, molecular dynamics, free energy calculations)	
Webinar	2.1, 2.2, 2.5
Short seminars with questions and documentation on how to install specific software packages/codes (own environment & cloud deployment)	
Webinar / written documentation	3.6, 3.7, 4.4
An Entry level course, potentially based on modules PIs currently teach, topics to include: docking, molecular dynamics, free energy calculations (in collaboration with webinars)	
Online course	2.1, 2.2, 2.5
Mathematical basis of different modelling approaches	
Webinar	2.2, 2.3, 2.5
HPC for Life Scientists (based on material created for BioExcel training event 4&5)	
Online course	3.1-3-8 & 4.2-4.5
Virtual training on BioExcel pilot codes	
Virtual course	2.5, 2.6

Table 10: List of online and blended-learning resources that are under discussion for the second reporting period.

3.5   Long term impact

BioExcel routinely collects course feedback for all BioExcel training activities, the responses provide us with valuable information on how to improve our events and the direction of the training programme. However, these responses are generally collected on the last day of the event and therefore do no provide us with a lot of information about the long-term impact of the BioExcel training programme. We want to know if on a longer time scale attending our training courses has had a positive impact on their research/career and whether or not they have started using the tools that were introduced to them during the training event.

In order to find out that information BioExcel will be sending a long-term feedback survey to all delegates who have attended BioExcel training courses 6-12 months after the training event. The first set of information requests will be sent out in April 2017 (PM18), covering the first and second BioExcel training event.

We have based the template for our long-term feedback survey (see Appendix 7.4) on the EMBL-EBI survey. EMBL-EBI has been collecting structured long-term feedback for their training programme for all courses completed since 2015; for selected courses feedback has been collected for as far back as 2010. For EMBL-EBI the response rate to long-term survey is approximately 25%.

In addition to the survey we are currently discussing a potential collaborative effort with ELIXIR EXCELERATE to use interview with delegates as a way to analyse the long-term impact of our training activities. The BioExcel Summer School 2017 - Foundation skills for HPC in computational biomolecular research could be used for these interviews. The event is far enough into the future that we can approach delegates in advance to ask them if they would be willing to be involved in the impact research. If any delegates also attend the "Hands-on Introduction to HPC for Life Scientists" that would give us a unique opportunity to follow the learning process of an individual over a significant period of time.

3.6   BioExcel Event Guidelines

Here we outline the BioExcel dissemination and event guidelines. It is a "living document" curated by WP3/WP4 and suggestions are being added by anyone in the CoE. The document provides useful information about the available communication channels, standard working practice and tips & tricks. As of PM18 the content is added to the D4.4 and D4.5 deliverables but the document will evolve beyond that timeframe.

Over time we aim for this document to reflect the combined institutional knowledge and expertise of the BioExcel partners in organising dissemination and training events. It will be useful for new people joining the consortium or people that become involved in organising events without having a lot of previous experience.

It is related to BioExcel WP4 risks
R16: Low numbers of delegates at training events or workshops

The guidelines in section 3.6.1 are an extract of the full document and specifically relate to training events. Guidelines on BioExcel dissemination channels, images, webinar dissemination, and Interest Group events are presented in D4.4. The full document (at PM18) can be viewed https://docs.google.com/document/d/1hqzdsnJ_Dxv0DLpwE3M4otWI3xWDqz EARyOjfLB3SzI/edit?usp=sharing

3.6.1   Training events

Pre-course preparation
- Define learning outcomes for the course
- Define objectives for the course
- Clearly defined audience for the workshop, including any prerequisites
- As a guideline try and open the registration event up to 6 months before the date (programme can be in draft with fixed beginning and end times); register interest list can be implemented at an earlier stage.

Programme
- Address course expectations at the beginning of the course. Ideally with time allocated to introductions. This could be in group if you have lots of people attending.
- Gamestorming has useful activities that can be used to capture participants expectations and icebreakers http://gamestorming.com/category/icebreakers/
- Reserve time to fill in the feedback form and a wrap-up/conclusion session at the end
- If possible try and mix hands on and lecture rather than big blocks (not always possible)
  - People's attention span are short
- Add sufficient time for breaks to the programme, besides allowing the participants to network it also gives you a time buffer if a session overruns.
- If budget allows add a socialising event e.g. dinner, excursion, reception etc
- Consider using a Slack channel, Gitter, or twitter hashtag to encourage community building
- Consider gender balance when putting your speaker list together

Post-event
- Post course survey - should be filled in during the course
  - Contact Vera Matser for the surveymonkey template, please use the template as WP4 KPIs are dependent on the questions.
- De-brief with trainers and organisers to discuss participant feedback and their own impressions of the course
- Define improvements for the next iteration if applicable
- Post 6 months survey to ascertain long-term feedback
  - Done by WP4 as part of quality assessment & impact

Tips & Tricks
- Make trainers/facilitators life easier by giving participants 2 coloured post-it notes during hands-on session, to be place on the top of the monitor (sticking up). Green = done with task, Pink = I need help or You've lost me. Gives a quick overview of the status of the room and lowers the barrier for shy participants to ask for help.

- Worried that people are going to leave early on the last day? Add a highly regarded keynote speaker at the end.

4 Knowledge Resource Center and Competency viewer

4.1 Overview

We reported in D4.3 that due to technical difficulties with tagging and searching training resources using the BioExcel competencies, when using a wordpress plugin, we have opted for a phased implementation of the Knowledge resource center (also referred to as Knowledge Base).

The first phase of the Knowledge Resource Center was implemented in Q4 of 2016 and uses a google spreadsheet that allows filtering based on competency, domain and resource type (online vs face-to-face).

Phase two of the Knowledge Resource Center is part of a larger EMBL-EBI project aimed at integrating the concept of a Knowledge Resource Center with an interactive competency framework viewer. The competency profile creates a consistent definition system of competencies based on knowledge, skills and behaviours requirements that can be mapped onto learning resources. The larger project is funded by EMBL-EBI to benefit a number of Horizon2020 grants that use a competency-based training programme approach. We would like to release the work in an open-source manner so that it can benefit the wider competency framework community (for example: lifetrain), how to visualise a competency framework in a user-friendly way has been an ongoing challenge in this field.

4.2 Description of the larger Knowledge Base – Competency Viewer project

We want a user to be able to search the database of training resources based on topic, free text or event type (face-to-face or online). However, what will make the tool more powerful is if a user could also search via the BioExcel Competency Profile. By viewing the BioExcel competency profile a user can click on any of the competencies and see which training resources are available for the user to increase their competence in that specific domain.

Long-term we would like the user to be able to rate themselves on the BioExcel Competency Profile, e.g. 1 out of 5 for "Write his/her own scripts to perform tasks

in context of biomolecular research" and be able to make use of the recommended training resources and routinely re-assess their own development. An additional long-term feature would be if we can add reference personae to the system e.g. GROMACS developer and let a user compare themselves to the reference personae. This could be helpful if a user is considering a career as a developer and would like to assess which competencies to focus on.

An administrative interface will be needed to be able to update the competency profile and add additional training resources to the database when applicable.

The early work for this is part-funded by BioExcel (building of the Knowledge Resources Center prototype), while the remaining work is funded by EMBL-EBI.

BioExcel will be a use case for the extended Competency Viewer and will directly benefit from the extended functionality. We anticipate some user requirement gathering and user testing opportunities, which will be organised through the training interest group (see section 5.1). Organising these sessions through the BioExcel Training Interest Group will allow us to raise the profile of the training interest group, promote the BioExcel Training Programme as well as gather input for the Knowledge Resource Center and Competency viewer project

4.3   Prototype – Knowledge base

The first prototype contains the knowledge base aspect and is used to describe the specification for the API (figure 2 and 3).

Figure 2: Annotated screenshot of the prototype containing only the knowledge base aspect.

Figure 3: Screenshot of the prototype containing the knowledge base illustrating the free text search functionality.

Our current aim is to have the Knowledge Resource Center operational by the end of May 2017.

5   Collaboration

5.1   Training Interest Group

5.1.1   Aim of the Training IG

To aid collaboration with other training professionals we have started a Training Interest Group within BioExcel; the Interest Group (IG) could include trainers of specific resources (e.g. HADDOCK, GROMACS, CPMD etc.), lecturers and training coordinators. The Training Interest Group is the channel through which BioExcel collaborates with other training initiatives (e.g. other CoEs, PRACE) so that challenges common to the training community (e.g. limited resources, providing on-demand training, lack of time for continuing professional development (CPD), trainer recognition) can be addressed together rather than in isolation.

5.1.2   Members

The BioExcel Training IG currently has 20 members; this number is not very high yet. As such this is not surprising since we have not been able to engage with potential users as much as we had hoped. To date we have one completed webinar for the Interest Group, which took place on 23rd November 2016. (Video recording available on: http://bioexcel.eu/webinar-9-defining-training-requirements-for-

biomolecular-researchers-with-high-computational-needs-2016-11-23/).
Outlined in section 5.1.3 are the plans for 2017, we anticipate being able to grow the Interest Group much more efficiently during the rest of 2017 as we have more activities planned. The number of members in this interest group will always be a relatively poor measure of the impact. A better measure will be the number of other initiatives and projects we are able to reach out to. Once the Interest Group has grown more we will be trying to quantify this impact better, currently we are in touch with ELIXIR, ELIXIR UK, CompBioMed, PRACE and a range of universities and research institutes.

5.1.3   Future plans

- Webinar: Working title is "Engaging the hard to engage – how to get life scientists/medics to embrace HPC/HTC". Delayed to Q2 due to other priorities.
- Online discussion: there is a recognised need for better coordination of training needs for life scientist in HPC. The aim is to initially set up an online discussion between PRACE, ELIXIR, BioExcel and potentially CompBioMed to get an overview of what each of the initiatives are planning to avoid duplication and make the available opportunities more visible. Additional online discussion (or webinar) topics that are being considered are – Use of VMs and Dockers in Training, Community building, and Best Practice in Virtual Training. These topics would benefit from having a more established IG and are therefore placed on hold until a suitable time.
- Face-to-face events: We are currently planning two face-to-face events for the Training IG.
  - The Training IG is planning to contribute to the BioExcel community forum (planned for 22-23rd November 2017). We anticipate three main aims for this event. First we would like to engage with as many BioExcel users as possible to introduce them to the BioExcel competency framework to get additional feedback on the completeness of the framework and the different levels at which the competencies are required for the different user categories. The second aim would be to get user feedback on the Knowledge Base and the further development of the competency viewer (see section 4). The third and last aim is to interact with BioExcel users to get general feedback on the BioExcel training programme, to engage with other training professionals and to anticipate new training needs.
  - The BioExcel Training IG is organising a workshop during the Lifelong learning Conference (EMBL Conference – LLL17), on 5-7th December 2017 in Heidelberg, Germany. URL for the event https://www.embl.de/training/events/2017/LLL17-01/. We will be able to reach a different audience at this event compared to any of the other BioExcel events. We aim to engage with education professional and the competency community to discuss the challenge of how to visualise and get users to engage with a competency profile. In contrast to the November community forum we anticipate to be able to gain feedback on the Knowledge base and

Competency Viewer from the point of view of the educational professionals. This community will be interested in using the viewer for their own project and will have a very different view on the project compared to a user who is interested in finding relevant training material.

5.2 Other collaborations

Not all collaborative efforts run through the Training Interest Group, though we actively encourage our collaborators to join the Interest Group.

ELIXIR (www.elixir-europe.org) - unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research.
ELIXIR EXCELERATE (https://www.elixir-europe.org/excelerate) - It will support a pan-European training programme, anchored in national infrastructures, to increase bioinformatics capacity and competency.

- As mentioned in section 3.5 we will be collaborating with ELIXIR EXCELERATE on determining long-term feedback through surveys and interviews.

Common Workflow Language (CWL) (http://www.commonwl.org/) – CWL is a specification for describing analysis workflows and tools in a way that makes them portable and scalable across a variety of software and hardware environments, from workstations to cluster, cloud, and high performance computing (HPC) environments. CWL is designed to meet the needs of data-intensive science, such as Bioinformatics, Medical Imaging, Astronomy, Physics, and Chemistry.

- ELIXIR, ELIXIR EXCELERATE, EMBL-EBI and BioExcel are joining forces to help establish more and better training resources and documentations for CWL. BioExcel will be used as a case study and will organise two webinars.

PRACE Advanced Training Centres (http://www.prace-ri.eu/) - The PRACE Advanced Training Centres (PATCs) provide top-class education and training opportunities for computational scientists in Europe and are the primary source for PRACE training portal materials. The BioExcel partners BSC and EPCC are PATCs and can act as liaisons to PRACE.

- The Spring School 2017 - HPC for Life Sciences was jointly organised by BioExcel and PRACE (KTH). In addition, the Hands-on Introduction to HPC for Life Scientists will be organised jointly with PRACE (EPCC).

Collaborative Computational Project for Biomolecular Simulation (CCPBioSim) (http://www.ccpbiosim.ac.uk/) - United Kingdom-based collaborative initiative, bringing together chemists, physicist and chemical engineers as well as researchers from all branches of "molecule-oriented" biochemistry and biology.

CCPBioSim engages with early career researchers and the non expert through the provision of tutorials and workshops enabling them to become proficient and productive users of biomolecular simulation techniques.

- Initial discussion have taken place with the aim to make both CCPBioSim and BioExcel more aware of the activities both initiatives are currently engaging in. At this stage we actively promote each other's events pending opportunity for further collaboration.

6 Final remarks

The launch of the Knowledge Resource Center will be used to kick off a promotion drive of the BioExcel online resources, including the online webinar repository. Attendance at conferences, such as ISMB/ECCB will be used to showcase the online resources. This will be supported by more attention for these resources in our social media strategy.

An aspect that has been discussed internally in the CoE is the concern about how to keep user training sustainable and how it will scale when the CoE expands and the number of users of the BioExcel resources increases. One of the strategies discussed is to internally train more people, thereby expand the pool of trainers who can be called upon to participate in training activities. By having dedicated training professionals (training can be a part-time activity) embedded in the BioExcel partners, these trainers can act as translators or mediators between the developers and the end-user while remaining at the cutting edge of code development.

A similar strategy is implemented at EMBL-EBI where trainers are embedded in the service teams, often also working as curators or scientific officers. The EMBL-EBI training team coordinates the training programme and organises cross-services training courses (including requests for courses from external organisation). The User Training Working Group (UTWG) brings together these trainers and members of the training team to allow for sharing of best practise and horizon scanning. A train-the-trainer programme assists with capacity building both internally and in the wider bioinformatics training community. In the case of BioExcel, the Training Programme would coordinate across the different resources and the Training Interest Group would allow for sharing of best practice and to provide support for people moving into the training field.

7 Appendix

7.1 BioExcel Competency Profile

Competencies	Knowledge	Skills	Behaviours
1. Generic Competencies			
1.1 Function effectively in teams to accomplish a common goal.	- Understanding of the context of the persons in the team - Aware of cultural differences	- Communication - Conflict management - Time management	- Invites two-way communication; actively listens/pays attention; able to excite participation and commitment from others - Delivers on his/her actions and inspires this behaviour in others - Informs others of relevant information appropriately and on time
1.2 Comprehend and comply with professional, ethical, legal, security, and social issues and responsibilities, and uphold these in the workplace as appropriate.	- Broad understanding of the relevant ethical, legal, and social issues - Understanding of internal ethical and legal policies - Understanding of organisational responsibilities	- Can interpret the letter and spirit of policies (ethical, legal, etc.) - Applies risk management skills	- Confronts or reports potentially unethical behaviour or behaviour at odds with the organisation's values - Complies with internal policies - Respects confidentiality

1.3 Communicate effectively with a range of audiences, technical and nonprofessional.	- Broad understanding of communication models - Can distinguish between different target audiences and identify the most appropriate means of reaching them	- Active listening skills - Appreciates the value of social media and uses it effectively to communicate effectively with different audiences - Able to work effectively with creative professionals (e.g. designers, animators, journalists - Ability to translate complex concepts to different audiences	- Communicates clearly and precisely with people at all levels - Delivers points in a structured and logical manner - Seeks out openness from others; gives straightforward, candid opinions to all - Communicates effectively even in difficult situations - Anticipates audience needs and tailors communications to the audience and the context
1.4 Engage in continuing professional development (CPD).	- Aware of organisational changes and external environment	- Face-to-face communication skills - Ability to objectively review own performance	- Demonstrates curiosity to learn new skills or knowledge outside current expertise - Prioritises and plans for CPD
1.5 Identification and understanding of user needs.	- Broad understanding of scientific disciplines and its ecosystem - Recognises different types of user, their interests and motivation	- Ability to differentiate wants from needs - User requirements/needs collection - Expectations management	- Proactive in discovering different users and their motivation - Actively facilitates gathering of use cases and user requirements

	- Appreciates the importance of user experience		- Seeks out and acts on user feedback
2. Scientific Competencies			
2.1 Apply computing expertise appropriate to the discipline and level of expertise (software as well as hardware).	Knowledge including but not limited to: - Computer architecture - Essential Algorithms - Programming - Databases & data structure - Information security	- Evaluates software - Understands applicability, optimisation and scope of different tools	- Keeps up-to-date with knowledge domain
2.2 Apply expertise in formal & natural sciences appropriate to the discipline.	- Has a deep comprehension of biological problems - Comprehends that a biological theory is not necessarily true - Fundamental scientific knowledge (biology, chemistry, physics) - Mathematics and statistics	- Has an interdisciplinary view - Asks relevant, well-defined biological questions	- Takes a comprehensive approach to problems - Consults with colleagues - Looks for prior work - Displays scientific humbleness (own and other results) - Allows for unexpected results
2.3 Detailed comprehension of the scientific process and an ability to apply it (e.g. experimental design; inclusion of controls; QA, QC, processing, visualisation and interpretation of results).	- Knows where to find help - Comprehends the need for positive and negative controls - Comprehends the need for replicates (and what type, how many etc.) - Comprehends the principle garbage in, garbage out (GIGO) (for example generic or	- Formulates a hypothesis - Tests a hypothesis - Accurately judges the validity of his/her results	- Asks for help and/or verification - Prepares before starting a new process

	custom parameters for molecular dynamics) - Has a deep understanding of his/her own data		
2.4 Comprehension of, and compliance with, licensing policy.	- Is aware of the different types of licensing - Understands the significance and potential ambiguity of licensing depending on where you work (academic/industry, funding etc.) - Understands the difference between open and closed licenses	- Chooses appropriate license for his/her own software	- Always checks the license information for software - Questions ambiguity in license information - Checks software is licensed
2.5 Use a computer-based system, process, component, or program to meet desired needs in a biomolecular context.	- Knows where tools are available	- Chooses the right tool for the problem at hand	- Actively searches for available sources of support (training material, forum, helpdesk etc.)
2.6 Evaluate the ability of a computer-based system, process, component, or program to meet desired needs in a biomolecular context.	- Has a deep comprehension of biological problems - Comprehend the capabilities and limitations of computer-based system, process, component and programs.	- Creative thinking & problem solving - Has an interdisciplinary view	- Keeps up-to-date with emerging techniques and applications
2.7 Comprehension of the local and global impact of high-performance computing (HPC) and high-throughput computing (HTC) on	- Understand what areas are impacted by HPC technologies	- Accurately judges added value of HPC technologies for different scenarios and user groups	- Disseminate the importance of HPC technologies

individuals, organizations, and society.			
2.8 Comprehension of how data-driven science, data analysis and computational modelling can be combined to generate and test hypotheses (e.g. machine learning, data mining, pattern recognition).	- Understands that science is an iterative process - Comprehends that models require experimental validation	- Accurately judges the validity of his/her results	- Comprehends that results might be difficult to interpret
2.9 Identify and compile appropriate data sets to address specific research questions	- Has knowledge of different datasets - Is aware of the concept of unique identifiers - Comprehends the need to question / validate data	- Assesses data quality - Judges fit-for-purpose - Actively identifies new datasets - Tracks where data came from - Combines data from different sources - Data preparation/clean up	- Advertises the existence of good quality datasets to others - Contributes to public datasets when appropriate
2.10 Comprehension of, and compliance with, best practice in data management / organization / archiving and storage	- Knows and understands the data formats people are using - Is aware of the backup policy of the institution/compute resource - Ability to judge what data needs to be kept for storage	- Demonstrates knowledge of the file system structure - Tracks ever step in a process (traceability) - Version control - Backup - Documentation - Automates data analyses where appropriate	- Backs up his/her work in non-redundant ways - Documents his/her work - Has a version control system in place
2.11 Comprehension of, and compliance with, best practice in	- Comprehends the need for standards (including	- Annotates metadata (tools) - Uses ontologies	- Uses standards and ontologies

distributed data management and data management planning	ontologies) - Understands the importance of metadata - Comprehends the challenge surrounding legacy data (e.g. inaccessibility, resourcing, IP) - When to move data, and when to run code elsewhere	- Structure data, where applicable submit to databases - Safely and efficiently moves data	- Looks for existing ontologies rather than create his/her own ontology - Captures sufficient metadata - Adds his/her data in supplementary information of journal articles (e.g. SMILES, InChl for compounds) - Actively promotes the use of standards, metadata annotation
2.12 Search for, read and assess literature / manual	- Comprehend the need for a literature review (anything similar done? Data available?) - Is able to find relevant literature (different sources of) (e.g. PubMed, Europe PMC)	- Assessing quality (methods, results, etc.) - Critically read scientific literature	- Conducts a literature review when appropriate - Will look at sources of literature outside his/her comfort zone / discipline
2.13 Presenting your results to community (writing papers, conference presentation, YouTube)	- Knows which forums exist to present scientific results	- Write scientific journal paper - Write and present a conference talk - YouTube - Proficient at scientific poster design - Deep understanding of what to present and what not to present - Can target his/her	- Judge the appropriate medium to present through

		presentation at the right audience / level	
3. Generic Computing Competencies			
3.1 Comprehension of, and compliance with, good programming practice (as promoted, for example, by www.software-carpentry.org.	- Comprehends the need for best practice - Is aware of where (or with whom) to find examples of or written guidelines on best practice in his/her field	- Inspires others to adhere to best practice	- Actively promotes best practice by leading by example - Will try and identify best practice in a relevant area
3.2 Analyse a problem and identify and define the computing requirements appropriate to its solution (e.g., define algorithmic time and space complexities and hardware resources required to solve a problem).	- Know how to describe algorithmic performance and complexity (e.g. Big O notation)	- Is able to benchmark a system to get an estimate of the required computing time - Identify bottleneck in the system (network, CPU, RAM, etc.)	- Write documentation
3.3 Apply knowledge of the operating system.	- Efficiently navigates his/her way around the Operation System (OS) - Is transitioning from Graphical User Interface (GUI) to command line (if not already proficient) - Demonstrates knowledge of the file system structure - Knows where to find the location of important configuration files &	- Creates and manages files and directories in a system using a Graphical User Interface (GUI) and using command line - Connects to remote systems with a GUI and the command line - Be able to use Unix/Linux features like pipes & redirection - Changes access permissions	- When you can't do something, search for it on the web - Back-up files - Uses access permissions appropriately

	applications - Is aware of OS-related file format differences	when required - Creates and manages files and directories in a system	
3.4 Write/adapt computer programs (software development) for biomolecular simulations.	- Recognises common programming concepts like loops and function calls - Comprehends the pros and cons of different programming languages - Knowledge of existing code/libraries to re-use	- Adapt existing programs - Compile code - Run test - Recognises appropriate file types - Knows when quick and dirty is appropriate and when writing for reusability is appropriate - Program in an appropriate language - Writes portable code - Writes efficient code - Writes appropriate tests - Debug code	- Conform to best practices for writing reusable code - Uses revision control - Uses a text editor with support for programming
3.5 Write his/her own scripts to perform tasks in context of biomolecular research.	- Knowledge of existing command/libraries to reuse - Judges when a task should be automated	- Is able to automate the process of executing some processes remotely - Write & debug scripts	- Uses appropriate scripting language
3.6 Install biomolecular simulation software on his/her computer.	- Knows where to download software & dependencies to install - Comprehend the difference between kinds of binaries & source	- Select appropriately packaged code - Be able to use package managers - Be able to revert a system to a known state	- Reads the README - Checks licencing before installing/running software - Appreciates the impact of installing/running software

	- Awareness of different types of account (e.g. user, admin, etc.) and when they should be used - Deep understanding of his/her platform (hardware, OS version, ...)		on other users of the machine
3.7 Deploy and test non-commercial software, including software that is built collaboratively and on a volunteer basis.	- Git (repositories), GitHub, SVN, mercurial-versioning	- debugging (Mission Control, VisualVM,..)	- work in collaborative fashion
3.8 Apply knowledge of systems monitoring (e.g. queue monitoring, systems availability and optimisation, storage used; scheduling maintenance at appropriate times and communicating this to users).	- Remote administration tools	- Linux/UNIX sysadmin commands - LSF commands - Recognise bottlenecks in the system (Network, CPU, RAM, etc.) - Forward planning	- Actively consider the impact on the user
4. Parallel Computing Competencies			
4.1 Assess computational workflow systems and their potential benefits.	- Is aware that workflow systems exist and what they do - Comprehends the pros and cons of different workflow systems	- Evaluates and selects an appropriate workflow system - Install workflow manager - Ability to run a workflow - Write and modify a workflow	- Seeks out and makes use of appropriate existing workflows and workflow components
4.2 Apply knowledge of batch system	- Understands what resources are required per job - Demonstrates knowledge of	- Writes a good batch script making use of available batch system features	- Reads the documentation - Ask the right person for help

	the concept of queues and the runtime environment - Know about versions of software that are installed - Can estimate who is responsible for observed issues (a hardware issue? a software issue? a user environment issue?)		
4.3 Write computer programs that can run on a parallel computer	- Git (repositories), GitHub, SVN, Mecurial-versioning	- Programs using an appropriate language/paradigm - Recognises (independent) units of work and potential parallelism (and any dependencies) - Uses appropriate tools to debug, profile, refactor parallel code	- Reads the documentation
4.4 Assess advantages and limitations for deploying, executing and optimising computations in a cloud/grid/HPC environment	- Evaluates the pros and cons of running computations in the cloud - Evaluates the pros and cons of using grid technologies - Demonstrates knowledge of container technologies (e.g. Docker) - Demonstrates knowledge of	- Routinely undertakes a cost-benefit analysis to understand whether cloud resources are the best value - Able to package software/data and deploy to the cloud - Selects an appropriate grid/cloud provider	- Seeks out and makes use of appropriate existing Virtual Machines, etc.

	the concepts related to virtualisation	- Can build and provision a Virtual Machine (VM)	
4.5 Apply knowledge of performance profiling to measure suitability of computing platforms	- Know that there are tools to help measure performance - Know how your machine's architecture and hardware can affect the performance of your code	- Be able to profile a code - Recognise bottlenecks in a code - Be able to refactor code in order to remove bottlenecks	- Run codes using appropriate resources (number of processors, etc.)

Table 11: The BioExcel Competency Profile listing the competencies and the Knowledge, Skills and Behaviours needed to take full advantage of the services offered by BioExcel.

7.2 List of attendees per BioExcel training event

7.2.1 "BioExcel: addressing training needs for advanced simulations in biomolecular research", 3-4th May 2016, EMBL-EBI, UK

First Name	Last Name	Affiliation
Mark	Abraham	KTH
Alexandre	Bonvin	Utrecht University
Cath	Brooksbank	EMBL-EBI
Adam	Carter	EPCC, The University of Edinburgh
Mark	Forster	Syngenta
Emiliano	Ippoliti	Forschungszentrum Jülich
Ramyarooban	Kanapathywasam	Unilever
Predrag	Kukic	University of Cambridge
Lee	Larcombe	ELIXIR UK University of Edinburgh
Caroline	Low	Imperial College London
Lok Hin	Lui	University College London
Vera	Matser	EMBL-EBI
Steven	Newhouse	EMBL-EBI
Ilias	Soumpasis	Unilever
Andrea	Spitaleri	Istituto Italiano di Tecnologia
Marta	Strumillo	EMBL-EBI
Mikael	Trellet	Utrecht University
Sameer	Velankar	EMBL-EBI
Gitanjali	Yadav	National Inst of Plant Genome Research

Table 12: List of speakers and delegates for "BioExcel: addressing training needs for advanced simulations in biomolecular research", 3-4th May 2016, EMBL-EBI, UK.

7.2.2 "BioExcel: Workflow Training for Computational Biomolecular Research", 20–21 October 2016, Barcelona Supercomputing Center (BSC), Spain

Type	First Name	Last Name	Affiliation
Delegate	Adrià	Aterido	Rheumatology Research Grup, Vall d'Hebron Research Institute
Delegate	Christoph	Bünemann	Shire
Delegate	Esther	Camacho Cano	Centro de Biología Molecular Severo Ochoa
Delegate	Jordi	Camps	CNAG
Delegate	Maria	Chatzou	Centre for Genomic Regulation
Delegate	Giovanni	Cincilla	Molomics Biotech
Delegate	Antonio	Flores	Universidad de Málaga
Delegate	Jana	Gurinova	University of Vienna
Delegate	Felipe	Hernandes Coutinho	Universidade Federal do Rio de Janeiro
Delegate	Joan	Lopez de la Franca	Barcelona Supercomputing Center - Earth Sciencies Department
Delegate	Maciej	Majewski	University of Barcelona

Delegate	Domingo	Manubens	BSC
Delegate	Minos	Matsoukas	University of Patras
Delegate	Serena	Piticchio	University of Barcelona
Delegate	Moira	Rachman	Universitat de Barcelona
Delegate	Amrinder	Singh	CNAG/CRG
Delegate	Raul	Tonda	CNAG
Delegate	Andrea	Townsend-Nicholson	University College London
Speaker	Rossen	Apostolov	KTH Royal Institute of Technology
Speaker	Michael	Crusoe	Common Workflow Language project
Speaker	Paolo	di Tommaso	CRG - Centre for Genomic Regulation
Speaker	Daniela	Digles	University of Vienna
Speaker	antonio	espinosa	Universitat Autonoma de Barcelona
Speaker	Daniele	Lezzi	BSC
Speaker	Vera	Matser	EMBL-EBI

Speaker	Anna	Montras	Institute for Research in Biomedicine (IRB Barcelona)
Speaker	Stian	Soiland-Reyes	University of Manchester
Speaker	Gonzalo	Vera	CRAG
Speaker	Mark	Abraham	KTH Royal Institute of Technology
Speaker	Rosa M	Badia	BSC
Speaker	Darren	White	EPCC - University of Edinburgh

Table 13: List of speakers and delegates for "BioExcel: Workflow Training for Computational Biomolecular Research", 20–21 October 2016, Barcelona Supercomputing Center (BSC), Spain.

7.2.3 "BioExcel/PRACE Spring School 2017 – HPC for Life Sciences", 10–13 April 2017, KTH, Stockholm, Sweden.

Type	First Name	Last Name	Affiliation
Delegate	Mark	Abraham	KTH
Delegate	Alma	Andersson	SciLifeLab
Delegate	Cathrine	Bergh	KTH
Delegate	Michal	Biler	KTH
Delegate	Joseph	Brock	Karolinska Institute
Delegate	Sonali	Chavan	University of Gothenburg
Delegate	Xiaoyu	Chen	KTH
Delegate	Mahsa	Ebadi	Uppsala University
Delegate	Karim	Elgammal	KTH
Delegate	Jerry	Eriksson	HPC2N, Umeå Universitet
Delegate	Stefan	Fleischmann	DBB, Stockholm University

Delegate	Arno	Foerster	Goethe Universitaet Frankfurt
Delegate	Federico	Gallino	SAEA Getter Spa
Delegate	Inacrist	Geronimo	Swedish University of Agricultural Sciences
Delegate	Mohammad Mehdi	Ghahremanpour	Uppsala University
Delegate	Arina	Gromova	SciLifeLab Stockholm
Delegate	Fredrik	Grote	Stockholm University
Delegate	Eva	Hellsberg	University of Vienna
Delegate	Michael	Holmboe	Umeå Universitet
Delegate	Mariama	Jaiteh	Uppsala University
Delegate	Stefanie	Kampen	Uppsala University
Delegate	Balamurugan	Kanagasabai	KTH
Delegate	Stefanie	Kickinger	University of Vienna
Delegate	Vladyslav	Kravchuk	Taras Shevchenko National University
Delegate	Junhao	Li	KTH-Royal Institute of Technology

Delegate	Laura	Llorach Pares	University of Barcelona - Mind the byte
Delegate	Magnus	Lundborg	SciLifeLab
Delegate	Marie	Lycksell	KTH chemistry
Delegate	Mehdi	Mahmoodinia	Norwegian University of Science and Technology
Delegate	Magdalena	Majekova	Slovak Academy of Sciences
Delegate	Pierre	Matricon	Uppsala University
Delegate	Amina	Mirsakiyeva	KTH
Delegate	Anne-Elisabeth	MOLZA	Aalto university
Delegate	Saeed	Mortezazadeh	IBPC-LBT Institut de Biologie Physico-Chimique
Delegate	Laura	Orellana	SciLifeLab
Delegate	Manoj Kumar	Panwar	Indian Institute of Technology Madras
Delegate	Anna	Popinako	Leninsky pr. 33
Delegate	Danial	Pourjafar-Dehkordi	Technische Universität München
Delegate	Szilárd	Páll	SciLifeLab

Delegate	Vishnu	Raju	KTH
Delegate	Stefano	Rebughini	SAEA Getter Spa
Delegate	Jaime Axel	Rosal	PDC / KTH school of biotechnology
Delegate	Ewelina	Rutkowska	University of Warsaw
Delegate	Karno	Schwinn	Goethe University Frankfurt
Delegate	Lenka	Stejskal	Birkbeck College, University of London
Delegate	Ning	Tang	University of Copenhagen
Delegate	YING	WANG	KTH
Delegate	Christian	Wennberg	SciLifeLab
Delegate	Annie	Westerlund	SciLifeLab
Delegate	Guomin	Yang	KTH
Delegate	Alexandra	Zahradnikova	Slovak Academy of Sciences
Delegate	Shaoqi	Zhan	KTH
Delegate	Yang	Zhou	KTH

Delegate	Andrei	Zimin	KTH
Speaker	Thomas	Cheatham	University Information Technology, University of Utah
Speaker	Jim	Phillips	University of Illinois, Urbana-Champaign
Speaker	Erik	Lindahl	Stockholm University
Speaker	John	Stone	University of Illinois, Urbana-Champaign
Speaker	Henric	Zazzi	KTH
Speaker	Michaela	Barth	KTH
Speaker	Vera	Matser	EMBL-EBI
Teaching Assistant	Pedro	Ojeda-May	Stockholm University
Teaching Assistant	Thor	Wikfeldt	KTH
Teaching Assistant	Cristian	Cira	KTH
Teaching Assistant	Mahan	Tourkaman	KTH

Table 14: List of speakers and delegates for "BioExcel/PRACE Spring School 2017 – HPC for Life Sciences", 10–13 April 2017, KTH, Stockholm, Sweden.

7.3 BioExcel Travel Grant

**Blog posts from the beneficiaries of the BioExcel Travel Grant**
For each of our BioExcel training events we make a limited number of travel bursaries available. We ask the beneficiaries of the BioExcel Travel Grant to write a blog post for us, either about their experiences at the course, their research interests (in layman's terms) or a relevant scientific topic.

THE POWER OF KNOWLEDGE
By Jana Gurinova
jana.gurinova@univie.ac.at
University of Vienna, Department of Pharmaceutical Chemistry
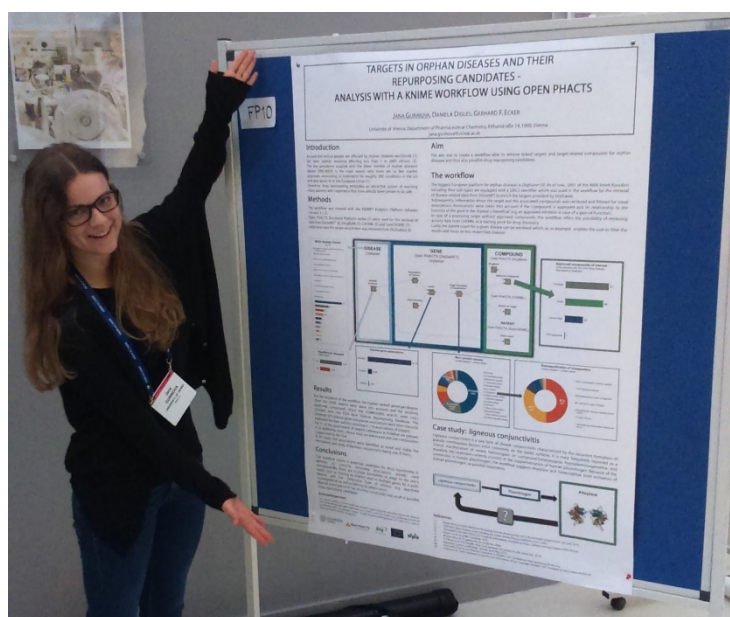
During my childhood I believed without a doubt that it is possible to know everything there is to know about all the disciplines that have been studied in the course of humanity. And even though I grew out of this belief, the reasoning behind it has never left me. Only when we know everything will we be able to make correct decisions, predict outcomes, and create something that is more than the sum of its parts. The same way a book is more than an arrangement of letters, or the human brain is more than a bunch of molecules.

A few centuries ago, the notion of knowing everything was not as far-away as nowadays, and the scholars of those times profoundly influenced our sciences. Just think of Isaac Newton or Leonardo Da Vinci. A portion of Da Vinci's work was devoted to the invention of a machine capable of flying. However, even though he came close, none of his inventions were actually able to fly, simply because although he did know everything there was to know at the time, the sciences themselves did not know enough. Flying is a privilege of our time, where disciplines have split, developed separately, and then came together again to make a machine weighing tens of tons able to fly.

When it comes to drug discovery, we were not so fortunate. The disciplines did split, and they did develop separately, but somehow they never came back to create this one thing that supersedes them all. A prediction on how molecules will act in the human body.

The more I get to know other disciplines, the more I feel like we truly do not know what our peers are capable of. Knowledge about the capabilities of today's science

is infinitely more powerful than complete knowledge of a single one. The moment you know who to ask and where to look, the impossible becomes reality.

In my case, the impossible became attainable by unifying the work of several disciplines, resulting in a workflow capable of predicting compounds which might be repurposed for the treatment of orphan diseases. The picture shows me presenting a poster of this work at the YMCS in Manchester. But as always, the method is not fool proof, which is why I am eager to get to know many more disciplines to ultimately create a prediction that supersedes its parts and truly creates value.

By Antonio Flores
Universidad de Málaga

The past 20th and 21st of October, I participated in the workshop organized by BioExcel "Workflow Training for Computational Biomolecular Research", they were two lovely days.

The organizers made us feel at home and we could drink coffee on demand :). Barcelona is an amazing city and the place of the workshop was also very nice, next to the supercomputer "Marenostrum" that we visited; it´s awesome. The only problem is that two days is not enough time to see the city.

The course content was very interesting, although my prior knowledge about workflow managers was very limited, the experts in each manager made the learning and use of programs very easy and I discovered some workflow managers that I could use in the future.

For me, the most interesting part of the workshop was the opportunity to talk with high level experts with whom I could discuss some of my research problems and they helped me to solve many of the doubts that I had before the workshop. I also had a chance to meet other researchers with whom to share concerns and strategies in addition to the possibility of continuing the contact with these people. I hope soon new workshops or courses are organized for BioExcel as interesting as this. No doubt, I would repeat.

By Esther Camacho
Centro de Biología Molecular Severo Ochoa

The BioExcel: workflow training for computational biomolecular research course held in Barcelona was a great meeting which allowed me to gain a better understanding of the utility of working with workflow managers when you have to deal with big data. I am currently working in a Next Generation Sequencing Core and I was surprised to see that many other colleagues of the field use workflow managers in their labs. I also learned about a kind of workflow managers which I didn't know (workflow managers without interface) and find them very interesting for my field.

I particularly enjoyed the visit to the Marenostrum supercomputer, one of the highest computer in Europe and the group dinner at La Traviata restaurant where I had the opportunity to speak closely with some of the speakers and attendants of the course. That was very important for a junior analyst like me.

The closing event Bring your own workflow was excellent for discussing our daily job problems and how workflow managers could help us. Here, the experienced speakers gave advice to the attendants about how to deal with these problems. In that sense, the speakers were very helpful and cooperative with us all the time. In the last part, we divided in groups according to our field, so we could make a deeper discussion about the pros and the cons of the different workflow managers and current limitations.

Before attending the meeting, I was concerned that the course would be orientated mainly to proteomics but it was not the case. The course had a good balance between proteomics, genomics and drug design. So, I think all the participants benefitted from the course without feeling that their field was not well-covered.

I am very grateful to have had the opportunity to attend the course. I have met excellent people that were, as well, experienced colleagues of my field. The course has fulfilled all my expectations.

7.4   Template for the BioExcel Long-term Impact Survey

BioExcel Long-term feedback survey

Feedback survey

As part of the development of our ongoing training programme for BioExcel we are interested in finding out how trainees have used the skills and knowledge they gained through participating in a course once they have returned to their study / work etc. As a recent participant of one of our courses we would like to know your opinions on the impact the course has had on you and your work.

Where possible, please take the time to provide written answers and explain as much to us as you can.

BioExcel Long-term feedback survey

BioExcel course details

This section asks about your experience of the course.

* 1. Which BioExcel course did you participate in?

2. Were you working in the field when you joined the course?

○ Yes

○ No

○ Other (please specify)

[   ]

3. Were you aware of BioExcel prior to the course?

○ Yes

○ No

○ Other (please specify)

[   ]

BioExcel Long-term feedback survey

After the course

This section asks about your experiences after the course.

* 4. During your course you will have learnt and used a variety of methods and resources. Since the course have you applied the following in your own research?

☐ Workflow managers

☐ Molecular dynamics codes

☐ Use of other tools

Please comment on specific examples

[   ]

* 5. Please comment on the impact that the course has had on your research and the research of others.

* 6. Did you establish any collaborations with other participants / BioExcel researchers during this course?

○ Yes

○ No

Comments

* 7. How many others have you taught the skills and/or knowledge that you learnt during the course this year?

○ None

○ 1-5

○ 6-20

○ 20+

Other (please specify)

* 8. Please comment on the best aspect of attending a BioExcel course.

* 9. Please comment on the worst aspect of attending a BioExcel course.

* 10. Would you recommend this course to others?

○ Yes, I already have

○ Yes

○ Maybe

○ No

○ Other (please specify)

BioExcel Long-term feedback
survey

Final thoughts

11. Please tell us any final thoughts you have about your experience of BioExcel training courses.

BioExcel Long-term feedback
survey

Thank you

Thank you on behalf of BioExcel CoE for taking the time to complete this survey. It is your feedback that helps us to improve and maintain high training standards.