Defining training requirements for biomolecular researchers with high computational needs

BioExcel Educational Webinar Series #9

Presenters: Cath Brooksbank
Host: Adam Carter

23 November, 2016

Partners

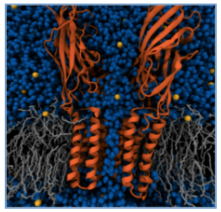
Funding

Horizon 2020
European Union Funding
for Research & Innovation

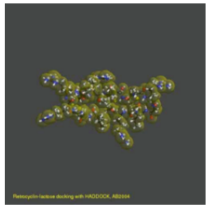
This webinar is being recorded

BioExcel Overview

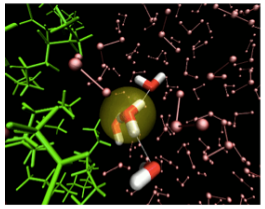
- **Excellence in Biomolecular Software**
  - Improve the performance, efficiency and scalability of key codes



MD simulations
/GROMACS/

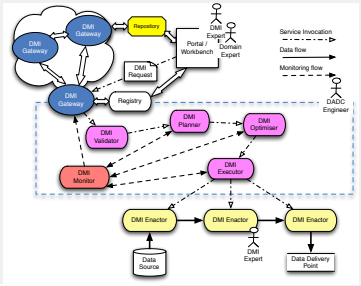


Docking
/HADDOCK/



QM/MM
/CPMD/

- **Excellence in Usability**
  - Devise efficient workflow environments with associated data integration



Key Workflows and Platforms

Galaxy PROJECT · Open PHACTS · COMPSs · Open for Innovation KNIME · Apache Taverna

- **Excellence in Consultancy and Training**
  - Promote best practices and train end users

Interest Groups

- Integrative Modeling IG
- Free Energy Calculations IG
- Best practices for performance tuning IG
- Hybrid methods for biomolecular systems IG
- Biomolecular simulations entry level users IG
- Practical applications for industry IG
- **Training**
- Workflows

Support platforms

http://bioexcel.eu/contact


Forums

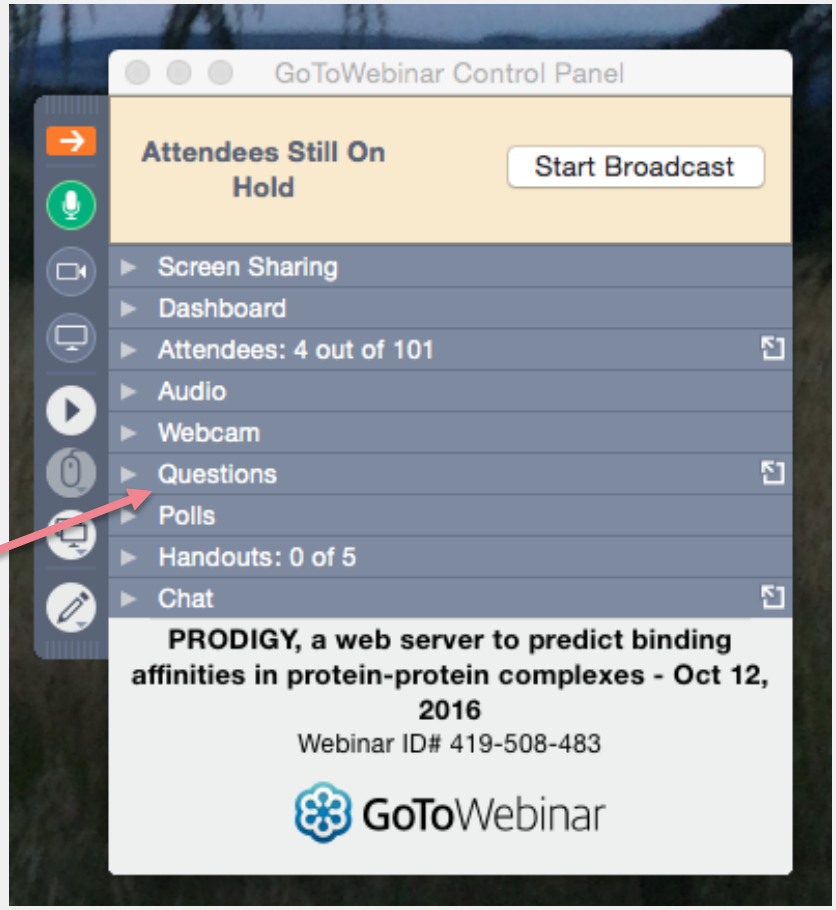
de Repositories
Cha nnel

Video nnel


Audience Q&A session

Please use the **Questions** function in GoToWebinar application

Any other questions or points to discuss after the live webinar? Join the discussion the discussion at http://ask.bioexcel.eu.



Today's Presenter

Cath Brooksbank

<presenter's slides here>

Objectives for this webinar

- Gain an overview of BioExcel and its goals
- Appreciate how to perform a simple training needs analysis as a precursor to developing a new training programme
- Discover how you can provide input into BioExcel's training programme
- Interact with others who have an interest in training for molecular life scientists, to learn from them and share your experience with them

Poll 1

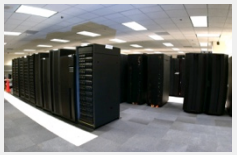
What is your motivation for joining this webinar?

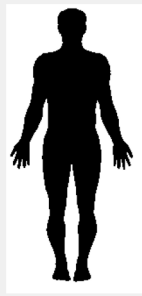
- To learn about BioExcel in general
- To learn about the BioExcel Training Interest Group
- To learn about our approach to performing a training needs analysis
- Other (please tell us using the chat function)

What is BioExcel?

Life science is becoming computationally intensive

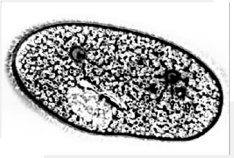
Population



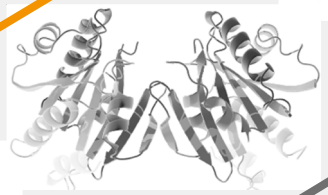


Organ

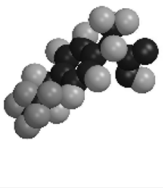
Tissue



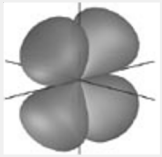
Cell



Macromolecules



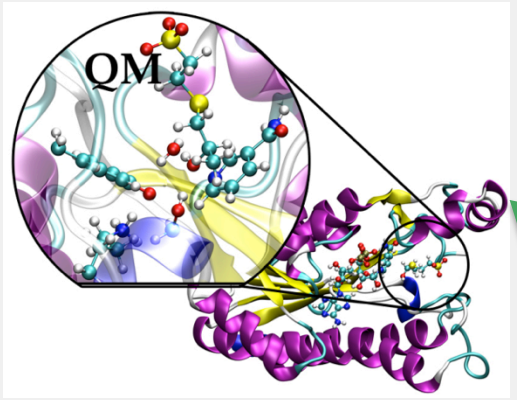
Small Molecule



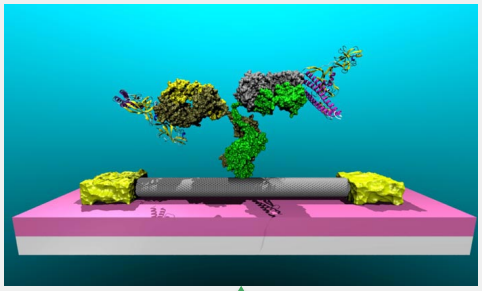
Atom

Biomolecular Modelling and Simulations

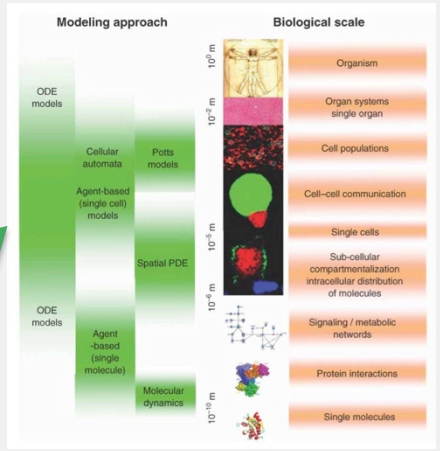
Electronic structure



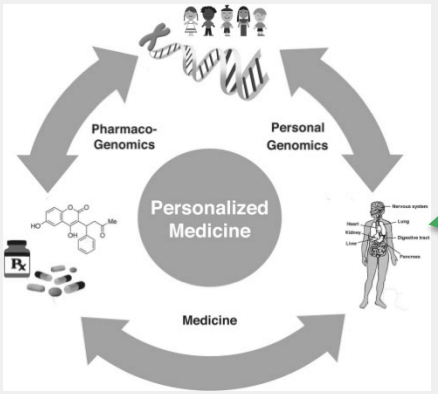
Biomarkers design



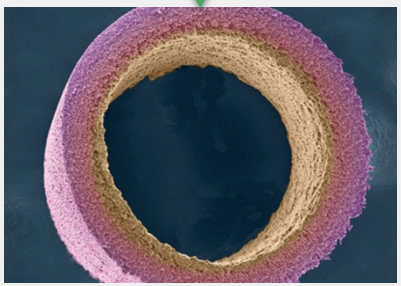
Physiology



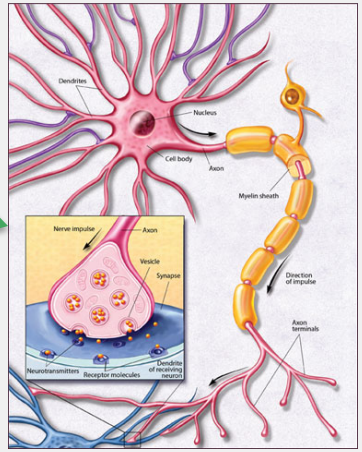
Personalized medicine



Biomaterials science and nanotechnology



Neuroinformatics



BioExcel Center of Excellence

SOLUTIONS

SOFTWARE
Widely used, fast and scalable codes for integrative modelling and molecular simulations

CORE DEVELOPERS
The scientists who wrote the code and know it best work with us!

WORKFLOWS
User-friendly and efficient systems for workflow executions and data processing

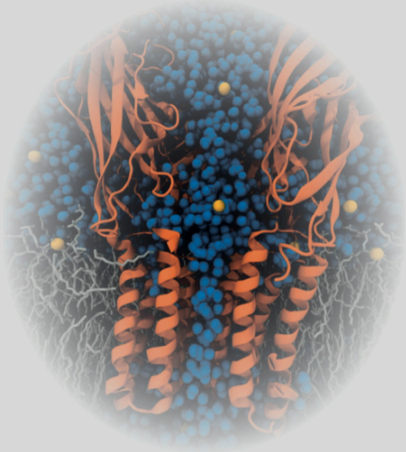
SERVICES

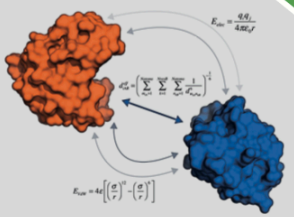
TRAINING
Webinars, "ask-me-anything" sessions, hands-on workshops for everyone from newbies to advanced users

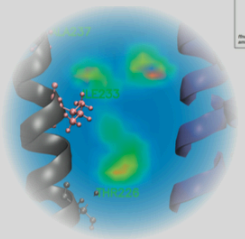
CUSTOMISATION
Tailored solutions adapted to your needs

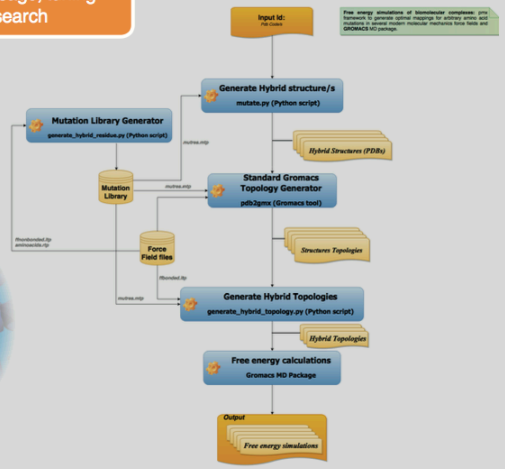
CONSULTANCY
Personalized support with software usage, tuning and scientific aspects of the research

ACADEMIA
INDUSTRY









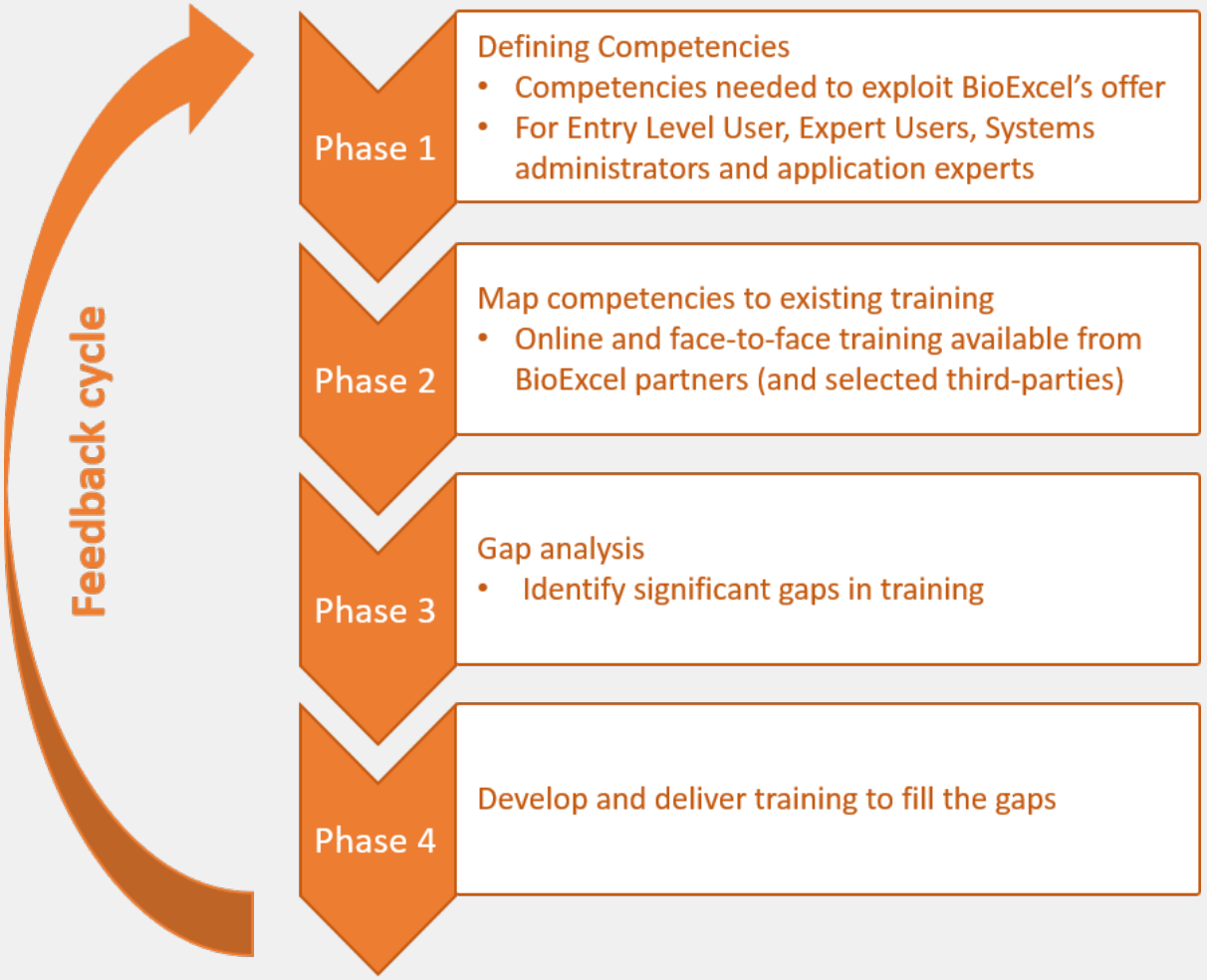
How are we defining BioExcel's training requirements?

Poll 2

What's your professional background?

- Life sciences
- Computer science
- Learning and development
- Other (please tell us using the chat function)

BioExcel Training Programme



What is a Competency?

Competency is ‘an observable ability of any professional, integrating multiple components such as knowledge, skills, values and attitudes’.

Acquisition can be validated objectively.

Shared ‘currency’ applicable to learning of all types and at all career stages

Competency profile

Defines the competencies required to fulfil a particular role

Typically defined by professional bodies / learned societies in collaboration with employers

An example…

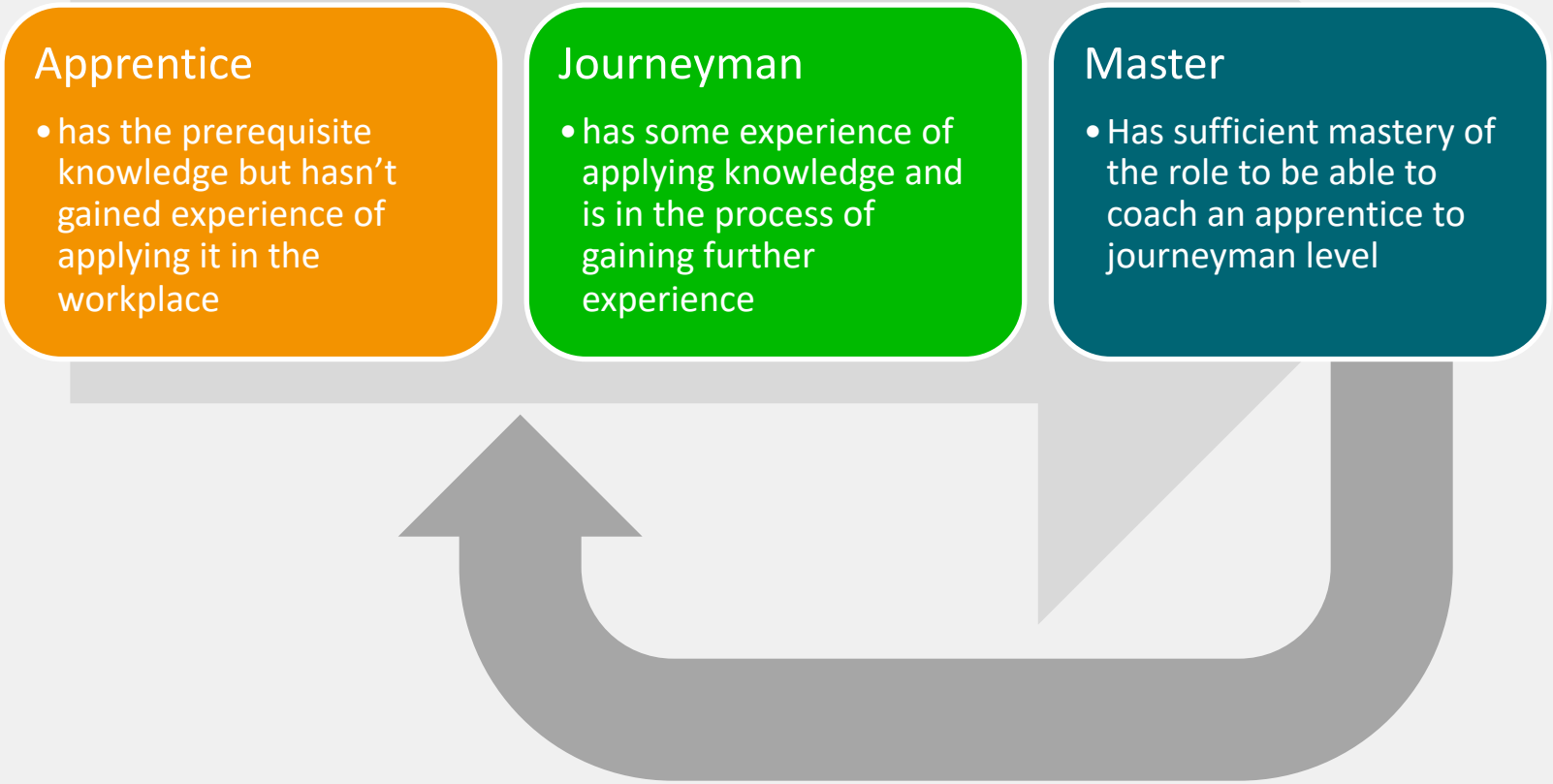
"Write his/her own scripts to perform tasks in context of biomolecular research"

Knowledge	Skills	Behaviour
Knowledge of existing commands/libraries to re-use	Is able to automate the process of executing processes remotely	Uses appropriate scripting languages
Judges when a task should be automated	Write & debug scripts	

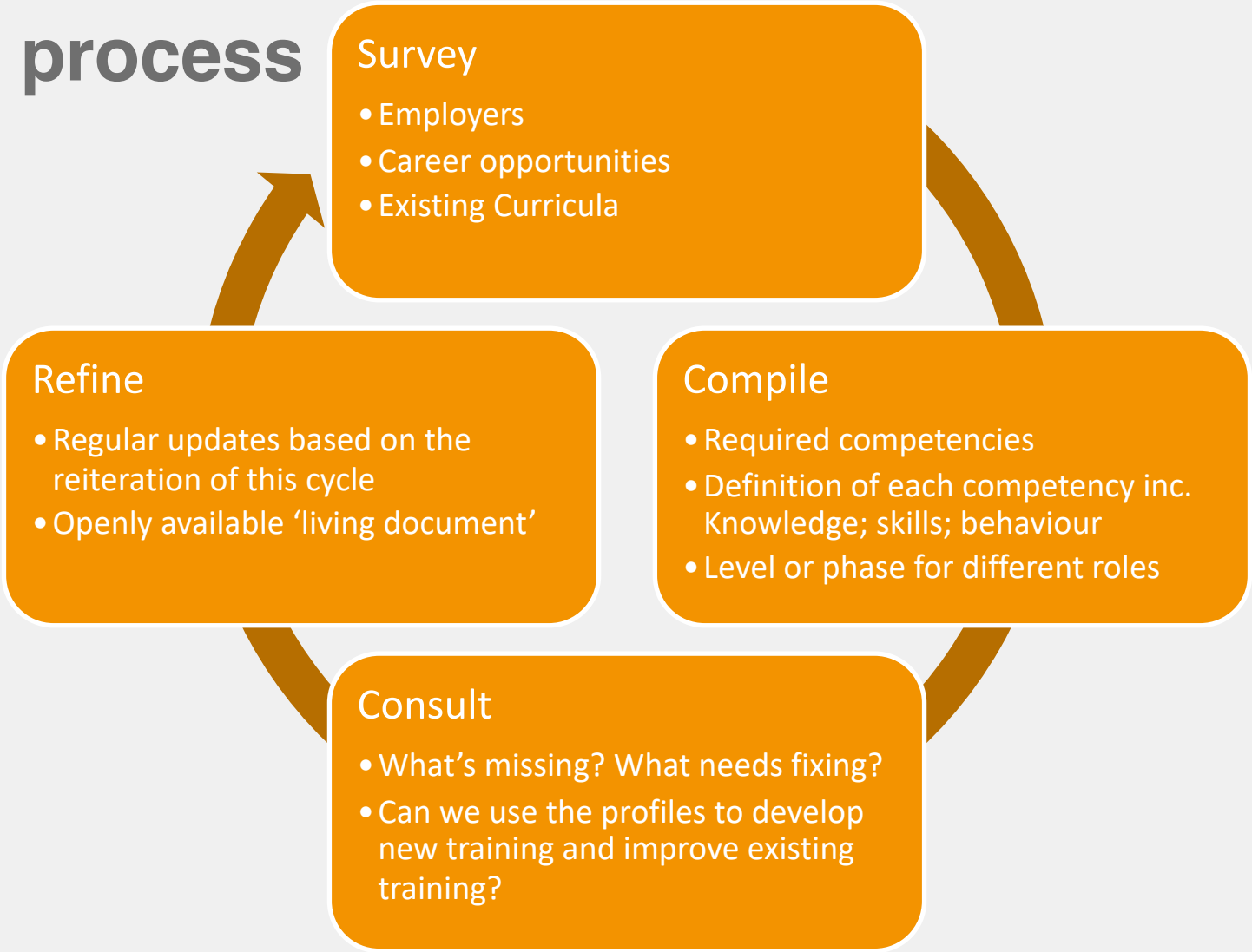
What level of competency is required?

N/A	• No competency in this area required
Awareness	• The professional appreciates what is possible in this area and how the area impacts on their own work
Working knowledge	• The professional has a firm underpinning knowledgebase in this area and applies it effectively in his or her day-to-day work
Specialist knowledge	• The professional actively contributes to advancement of the area, generating new understanding or new technology

How should their competency develop as they progress through their career?



The process


Survey
• Employers
• Career opportunities
• Existing Curricula
Compile
• Required competencies
• Definition of each competency inc. Knowledge; skills; behaviour
• Level or phase for different roles
Consult
• What's missing? What needs fixing?
• Can we use the profiles to develop new training and improve existing training?
Refine
• Regular updates based on the reiteration of this cycle
• Openly available 'living document'

BioExcel competency profile

Group related competencies together

- Generic competencies (5)
- Scientific competencies (13)
- Generic computing competencies (8)
- Parallel computing competencies (5)

Method and competency terminology have been aligned with similar international initiatives

Different types of users

- Entry-level user
- Specialist user
- Systems administrator / applications expert

→ You might be entry level for some competencies and specialist for others

Competency-based projects

RItrain
Driving Leadership for Research Infrastructures

LifeTrain
Driving lifelong learning for biomedical professionals

CORBEL
SHARED SERVICES FOR LIFE-SCIENCE

NHS
Health Education England

ISCB
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

bioexcel

Generic competencies (5)

Highlighted the most important ones for this group

Comprehend and comply with professional, ethical, legal, security and social issues and responsibilities

Function effectively in a team

Communication

Continuing Professional Development (CPD)

User needs

Scientific competencies (13)

Distributed data management & data management planning

Impact of HPC/HTC

Computing expertise

Expertise in formal & natural sciences

Licensing policy

Scientific process

Use computer-based systems

Evaluate computer-based systems

Find, read & assess literature

Data-driven science

Identify and compile datasets

Present your results

Data management

Generic Computing Competencies (8)

Operating system

Write/adapt programs

Good programming practice

Identify & define computing requirements

Install biomolecular software

SYSTEMS MONITORING

Write scripts

Deploy & test non-commercial software

Parallel computing competencies (5)

Workflow systems

Batch systems

Performance profiling

Parallel programming

Advantages & limitations for deploying, executing and optimising computations in cloud/grid environment

Want a closer look?

Competency survey

goo.gl/dEvs8N

Full profile

goo.gl/8OqKYK

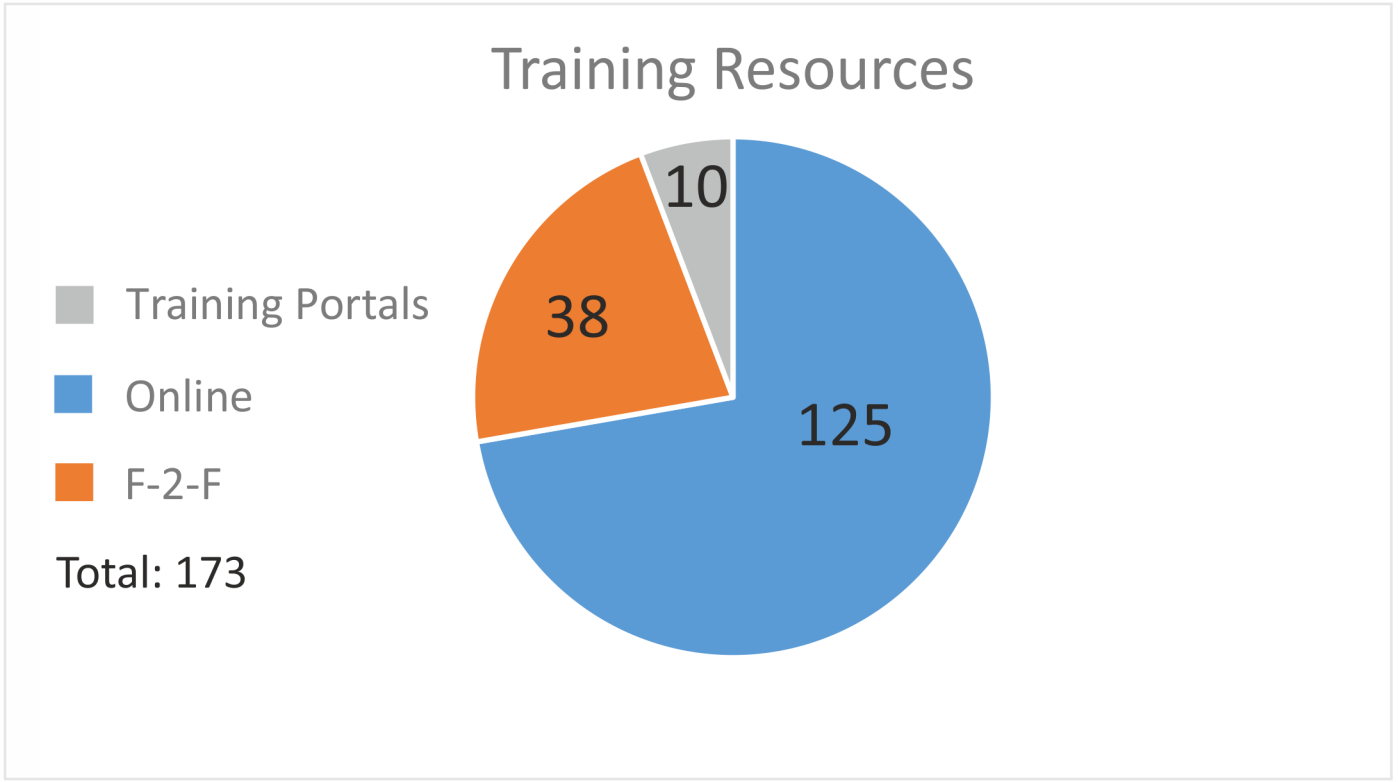
Let us know what you think!

What will this will be used for?

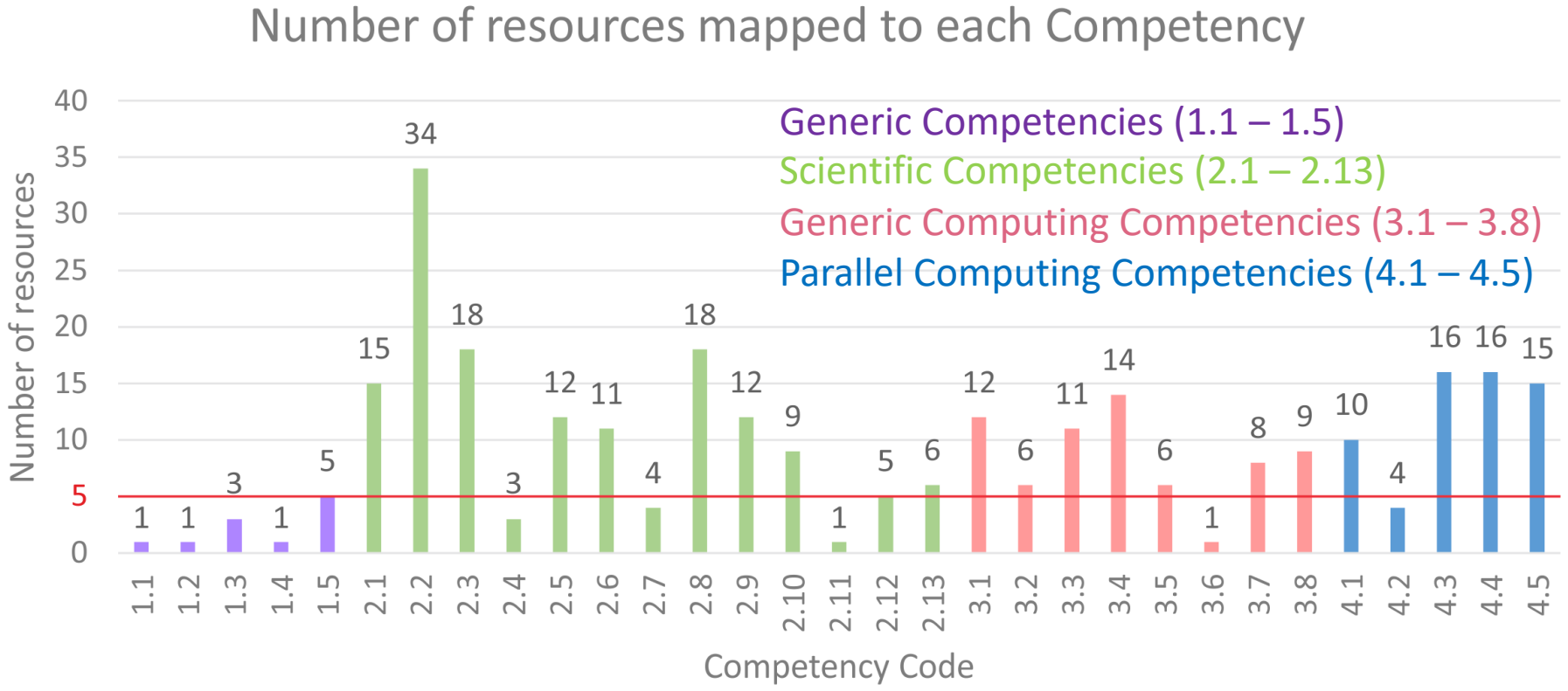
- **To develop high-value** training **materials** and adapt existing ones
- As components of **VM-based pre-packaged bundles** of applications and inputs for user-friendly exercises
- **To incorporate materials into workshops** at major Infrastructure events (e.g. PRACE, EGI, ELIXIR; INSTRUCT) or life science conferences (e.g. ECCB)
- **To collaborate with major related training activities** of PRACE, ELIXIR, CECAM, Pistoia Alliance and CCP-BioSim for reaching out to larger user communities

What kinds of training do we need to develop?

Identifying training resources

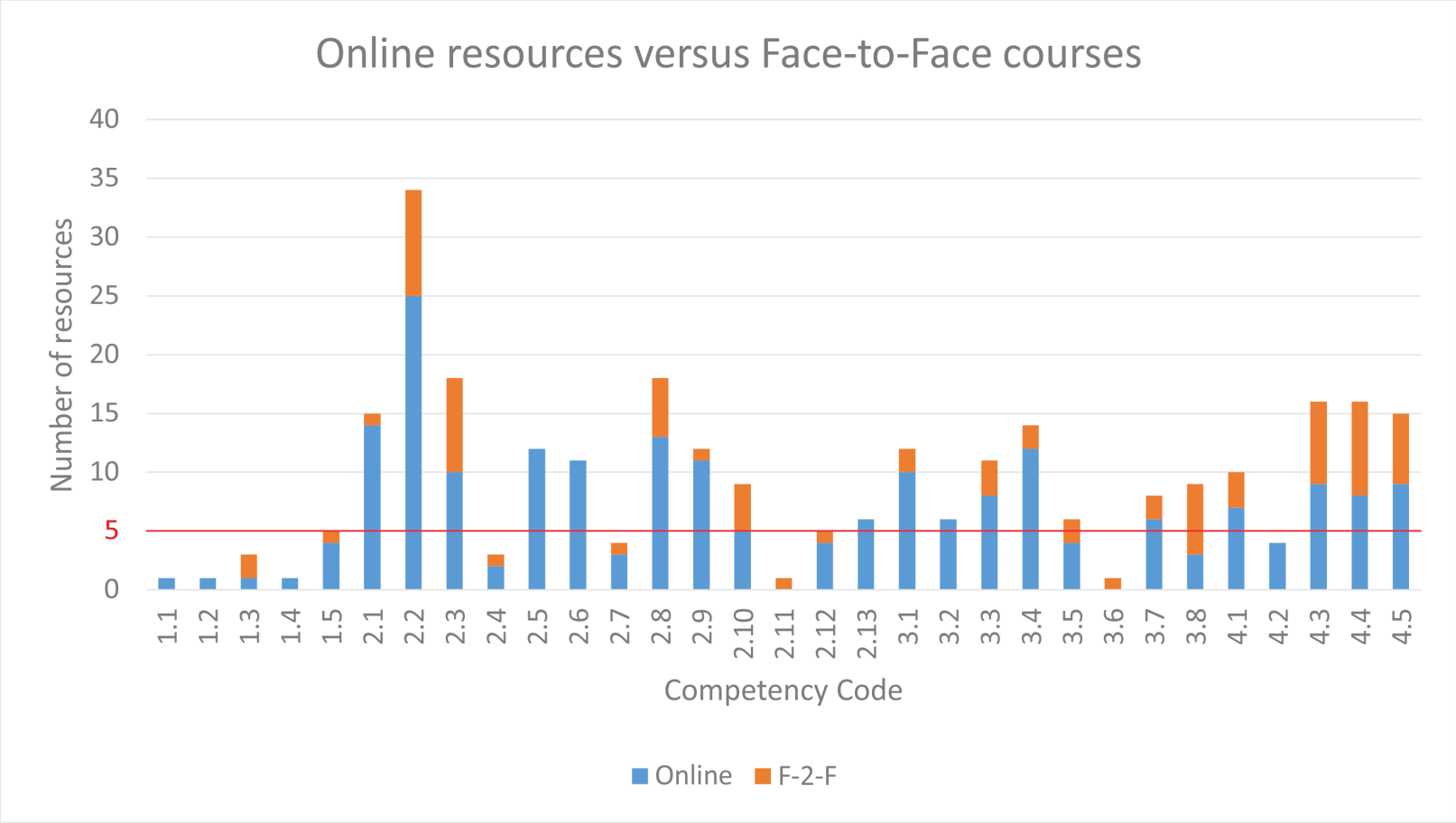


Mapping



5 (total training resources) picked as a minimum viable number for coverage

Online versus Face-to-Face



Gap Analysis

Key	Insufficient training resources identified but outside BioExcel scope	BioExcel or External training resources identified	Insufficient training resources identified

Side notes on Mapping & Gap Analysis

- Falsely optimistic view due to **partial mapping** of resources to competencies
- **Suitability** of the training resources for the BioExcel User Groups (entry-level users, expert users, and systems administrators)
    - Likely only internal or project associated resources

Gap Analysis – insufficient coverage

2. Scientific Competencies

2.4 Comprehension of, and compliance with, licensing policy.

2.7 Comprehension of the local and global impact of high-performance computing (HPC) and high-throughput computing (HTC) on individuals, organizations, and society.

2.11 Comprehension of, and compliance with, best practice in distributed data management and data management planning.

2.13 Presenting your results to the community (writing papers, conference presentation, YouTube).

Gap Analysis – insufficient coverage

3. Generic Computational Competencies

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).

3.3 Apply knowledge of the operating system.

3.6 Install biomolecular simulation software on his/her computer.

3.7 Deploy and test non-commercial software, including software that is built collaboratively and on a volunteer basis.

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).

Gap Analysis – insufficient coverage

4. Parallel Computing Competencies

4.2 Apply knowledge of batch system

BioExcel Training Programme

- **Entry Level Users:** Focus on bridging the gap between life-science and HPC/HTC computing (users transitioning from a bench-based molecular life science background).
- **Expert Users:** training needs for this group will probably be quite specific for each use case (e.g. specific software packages, optimisation). We anticipate input from the Interest Groups to inform highest priority training needs
- **Systems Administrators & Application experts:** Goals include Enabling **communication between this group and users in addition to** technical training

Poll 3

Which (if any) BioExcel user group do you identify most closely with?

Entry level user

Expert user

Sys admin/application expert

None (Please use chat to specify an alternative group)

BioExcel Training Programme

Face-to-Face training events

- **3-4th May 2016** - BioExcel: addressing training needs for advanced simulations in biomolecular research, EMBL-EBI
- **20-21st October 2016** - BioExcel: Workflow Training for Computational Biomolecular Research, BSC/IRB
- **10-13th April 2017** - PRACE & BioExcel Spring School – HPC in the Life Sciences (NAMD, AMBER, VMD, and GROMACS), KTH
- **3-7th July 2017** - BioExcel Summer School – Foundation skills for HPC in computational biomolecular research @EMBL-EBI
- **Autumn 2017/Early 2018** - Introduction to HPC for Life Scientists (with PRACE) @EPCC and/or BSC
- **Summer 2018** – Summer School 2018

BioExcel Training Programme

Sustainability

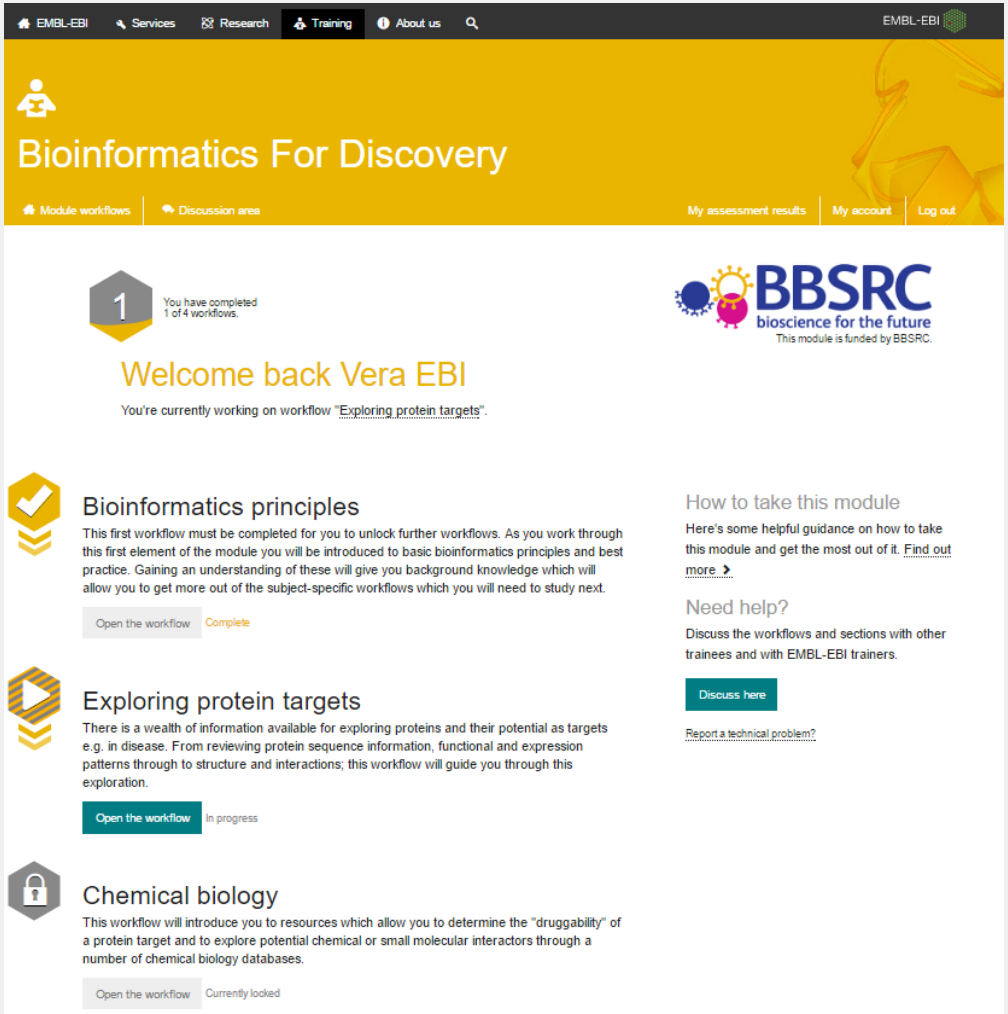
Webinars (in preparation)

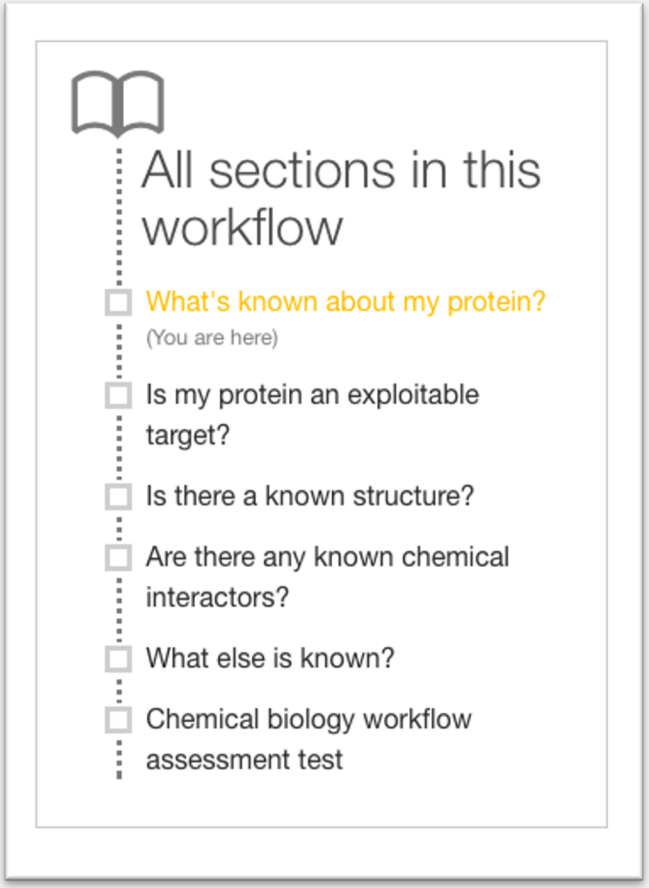
- Computational resources: local computer vs Cloud / HPC
- Quality control of structures / models
- Entry level webinars (Biomolecular modelling, Molecular Dynamics etc.)
    - Aimed at general science background
    - Turn into e-learning

E-Learning (workflow-based training)

- Repurpose course material from face-to-face
- Webinars
- Bespoke online courses
- Knowledge base

Workflow-based training





www.ebi.ac.uk/training/events/2017/bioinformatics-discovery-0

Collaborations (Training)

RItrain (http://ritrain.eu/) - mission is to improve and professionalize the training of managerial and leadership staff in research infrastructures (RIs).

CORBEL (http://www.corbel-project.eu/) - is an initiative of eleven new biological and medical research infrastructures (BMS RIs), includes competency-based training Work Package aimed at technical operators of BMS RIs.

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.

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.

CompBioMed (CoE) – (http://www.compbiomed.eu/) A Centre of Excellence in Computational Biomedicine

Training Interest Group (TIG)

Aims:

- Improve visibility of training initiatives and individual courses/resources. Members of the Training IG will be able to act as **liaisons** to other training initiatives
- Facilitate **collaboration** between projects to promote best practice and efficient use of resources
- Improve **communication** between "computational trainers" and "life science trainers"
- Promote the importance of **high quality training**

TIG is for you if you are responsible for, or have an interest in, training within your institute or project or are a training professional (trainer, coordinator, lecturer) in a life sciences and/or computational field

TIG - plans

Webinars

- 23.11.2016: Defining training requirements for biomolecular researchers with high computational needs (IG kick-off)
- Early 2017: Engaging the hard to engage – encouraging biomedical scientists to embrace high-end computing

Face-to-face

- Feb 2017: Using containers and VMs in training
- Jul 2017: ISMB/ECCB competency workshop – BioExcel use case
- 5–7 Dec 2017: Lifelong Learning in the Biomedical sciences (EMBL Conference)

What other activities would you like to see in the TIG?

Audience Q&A session

Please use the **Questions** function in GoToWebinar application

Any other questions or points to discuss after the live webinar? Join the discussion the discussion at http://ask.bioexcel.eu.