VIEWPOINT

Challenges for proteomics core facilities

Kathryn S. Lilley, Michael J. Deery and Laurent Gatto

Cambridge Centre for Proteomics, Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK

Many analytical techniques have been executed by core facilities established within academic, pharmaceutical and other industrial institutions. The centralization of such facilities ensures a level of expertise and hardware which often cannot be supported by individual laboratories. The establishment of a core facility thus makes the technology available for multiple researchers in the same institution. Often, the services within the core facility are also opened out to researchers from other institutions, frequently with a fee being levied for the service provided. In the 1990s, with the onset of the age of genomics, there was an abundance of DNA analysis facilities, many of which have since disappeared from institutions and are now available through commercial sources. Ten years on, as proteomics was beginning to be utilized by many researchers, this technology found itself an ideal candidate for being placed within a core facility. We discuss what in our view are the daily challenges of proteomics core facilities. We also examine the potential unmet needs of the proteomics core facility that may also be applicable to proteomics laboratories which do not function as core facilities.

Received: October 29, 2010
Revised: November 24, 2010
Accepted: November 29, 2010

Keywords:
Core facility / Proteomics standards / Standardization / Technology

1 Introduction

To date, many analytical techniques have been executed by core facilities setup by academic, pharmaceutical and other industrial institutions. Centralizing such facilities ensures a level of expertise and hardware, such as high-end mass spectrometers, which cannot be supported by individual laboratories, and is thus available for multiple researchers in the same institution. The cost of a high-end mass spectrometer can be well in excess of $500 000 and this excludes the necessary associated HPLC equipment. These costs are often above the reach of most individual research groups. The potential knock-on effect of this is that due to a lack of high-end, state-of-the-art equipment, few experienced mass spectrometrist/proteomic scientists will want to apply for vacancies at such facilities, which thus generates a lack of skilled staff. Often, the services within the core facility are also opened out to researchers from other institutions, frequently with a fee being levied for the service provided.

In the 1990s, with the onset of the age of genomics, there was an abundance of DNA analysis facilities, many of which have since disappeared from institutions and are now available through commercial sources. At the start of the 21st century proteomics technologies being utilized more and more by researchers, and the large capital expenditure required to purchase the necessary hardware, plus the high level of skill required to expedite proteomics technologies efficiently, resulted in this technology finding itself an ideal candidate for being placed within a core facility. Proteomics

Correspondence: Dr. Kathryn S. Lilley, Cambridge Centre for Proteomics, Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, Cambridge, CB2 1QR, UK
E-mail: k.s.lilley@bioc.cam.ac.uk
Fax: +441223760241

Abbreviations: ABRF, Association of Biomolecular Resource Facilities; **HUPO,** Human Proteome Organization; **LIMS,** Laboratory Information Management System; **QC,** quality control

technologies during this time were somewhat immature especially in comparison with similar transcriptomics facilities and it could be argued that the more successful proteomics core facilities were coupled with technology development. Irrespective of the shortcomings of technologies at this time, much investment was made by funding bodies, academic institutions and pharmaceutical companies to set up proteomics core facilities.

We discuss what in our view are the daily challenges of proteomics core facilities. We also examine the potential unmet needs of the proteomics core facility which may also be applicable to proteomics labs which do not function as core facilities and whose raison d'être is to develop technologies which are then applied to hand-picked projects.

2 Core facilities and their function

There is a wide spectrum of proteomics core facilities, from those solely involved in identifying proteins submitted as gel bands to those carrying out sophisticated quantitative measurements on large sample sets, to those whose technology development is driven by the needs of the researchers with whom they work. The majority have different levels of staffing with a manager in charge of operations supported by skilled technical staff. Periodically, proteomics core facilities are surveyed to 'take the pulse' of the modern facility. The Association of Biomolecular Resource Facilities (ABRF) is one organization which has conducted such studies, but to our knowledge an up to date comprehensive survey has not been carried out in recent times which aim to capture data from proteomics core facilities world wide.

There is significant variation in the parts of a process that the facility may be involved in; some may generate and/or process samples, others may have little to do with sample preparation. Some may give raw data to collaborators so that they can perform their own analysis, whereas others may characterize data sets and deliver a fully interpreted biological story to their colleagues. It is clear that many different levels exist in terms of a core team's involvement in the project design of experiments and how this varies across different facilities. A core facility may therefore require biochemists, statisticians, bioinformaticians, expert mass spectrometrists and above all, staff with excellent interpersonal skills who can manage expectation and be able to remain diplomatic and tactful when the ensuing data fail to deliver what the customer anticipated. It could be argued that the level of involvement of larger scale projects is crucial to their success, for example.

As much as the researcher needs to share vital detail with the core staff, what goes on inside many core facilities can easily been thought of as a black box as shown in Fig. 1. The samples are fed into the system, and a spreadsheet of data chugs its way out of the other side, hopefully in a format that can be cut and pasted into a dissertation or manuscript. Proteomics has many pitfalls and the researcher needs to be aware of these, before the project begins. The power of the technique and the biological variability of the samples to be processed need to be ascertained and discussed before a suitable experimental design can be planned to achieve meaningful data. The level of penetration into the proteome by any technique in terms of ability to measure within the concentration range that proteins of interest may occur and using a method suitable to the physico-chemical properties of the proteins of interest also needs much discussion at the onset of the project. Moreover, many core labs will have invested in a limited repertoire of technologies and hence may not be best equipped to carry out certain types of technologies. It is thus important to have educational tools on hand to assist in advising collaborators and also to have knowledge of other facilities which may be able to assist with methodologies not established within the local core lab.

It is vital that core facilities also keep abreast of technologies and enforce good quality controls (QC) to ensure that the often very sophisticated procedures are consistent.

3 Daily challenges of a core facility

There are many challenges which face proteomics core facilities. In addition to the ones already mentioned above, the facility must ensure the following are carefully considered:

(i) Tracking samples/data within the facility
(ii) Data storage
(iii) Keeping instrument downtime to a minimum
(iv) Instrument optimization
(v) Replacement of equipment with often costly state-of-the-art models
(vi) QC
(vii) Keeping abreast of new approaches technologies
(viii) Good management of time between research and service
(ix) Training and retention of staff
(x) Maintenance of funding.

The issues listed can all be major headaches for core labs. There are plenty of solutions available for sample tracking, usually in the form of sophisticated Laboratory Information Management System (LIMS) systems which are commercially available, but these often come at a price which a significant number of core facilities cannot justify or afford

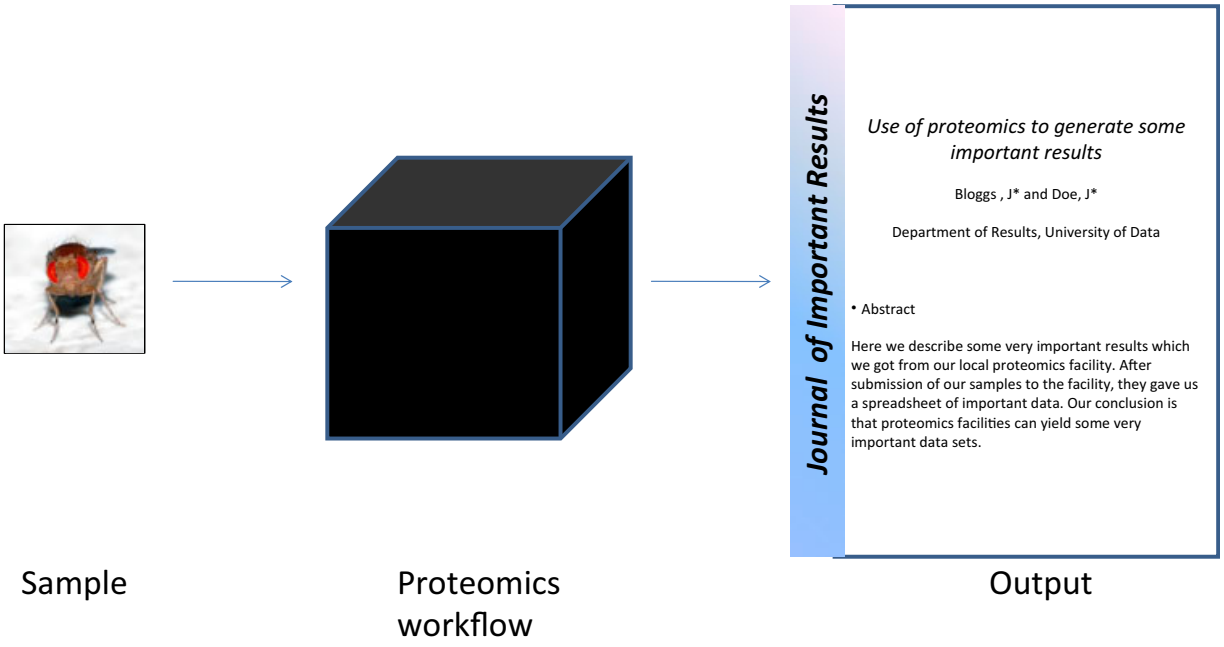

Figure 1. A common perception of proteomics workflow?

[1]. Sample tracking is not standardized in many facilities which may have home grown LIMS with variable degrees of sophistication. Many commercial LIMS systems may also not be applicable to the types of experiment undertaken in some core facilities.

Data storage can also be very challenging. Mass spectrometers create many terabytes of data per year, and with faster instruments collecting more data points per experiment, this problem continues to grow. There is the issue of what a facility stores, whether this be raw data or processed data, and for how long. There are many repositories now for storing data, for example PRoteomics IDEntifications database (PRIDE, http://www.ebi.ac.uk/pride) and Tranche (https://proteomecommons.org/tranche/), but usually deposition to these repositories occurs for data sets associated with published data. Many funding bodies now require data to be stored for a predetermined amount of time and also require data-sharing mechanisms to be in place before grant funding is released. The whole issue of depositing data with sufficient meta data for facile interpretation of the data is not trivial. Data storage comes at a price and adequate funding to ensure storage of data must be sought, otherwise proteomics data which itself is generally expensive to achieve may be inefficiently utilized.

Keeping instrumentation in good working order is one of the main challenges for any core facility. Instrument downtime results in delays in processing/analysing samples and if the facility is reasonably high throughput, this can lead to fairly large queues, which results in added pressure for the facility. Further pressure is then often applied from the client who was excitedly expecting their data days ago. The main measure which must be taken to ensure that downtime is kept to a minimum is good preventative maintenance, which should be the responsibility of the core facility's staff. However, as mass spectrometers become increasingly more complex in terms of electronic hardware, it becomes more and more necessary to take out the manufacturer's service contracts to ensure that any problems which occur and cannot be dealt with by in-house staff are swiftly rectified and also that the costs incurred throughout the year are covered by a one-off payment, thus making budgeting more simplistic. Of course, full service contracts are expensive and often beyond the means of many core facilities.

Instrument optimization is a perennial challenge for any analytical facility. Mass spectrometers must be calibrated, and their sensitivity checked regularly. Optimization of chromatography systems must also be established and depending on the types of protocols carried out in a facility, there are all manner of checks that need to be made to ensure reliable and reproducible protocols are being fully employed.

Every facility will have its own method of carrying out the above, more often than not using their favourite set of proteins/peptides for this purpose. Use of homemade standards sets may satisfy the need for internal standardization, but does not lend itself to cross laboratory standardization. Arguably sets of QC standards by which to test optimization are not well established in the field. If standard sets of well characterized, high QC standards were available, then any given facility would not only be able to monitor its own systems, but would be able to benchmark its capabilities against similar facilities elsewhere. For protein identification, facilities could monitor which proteins have been

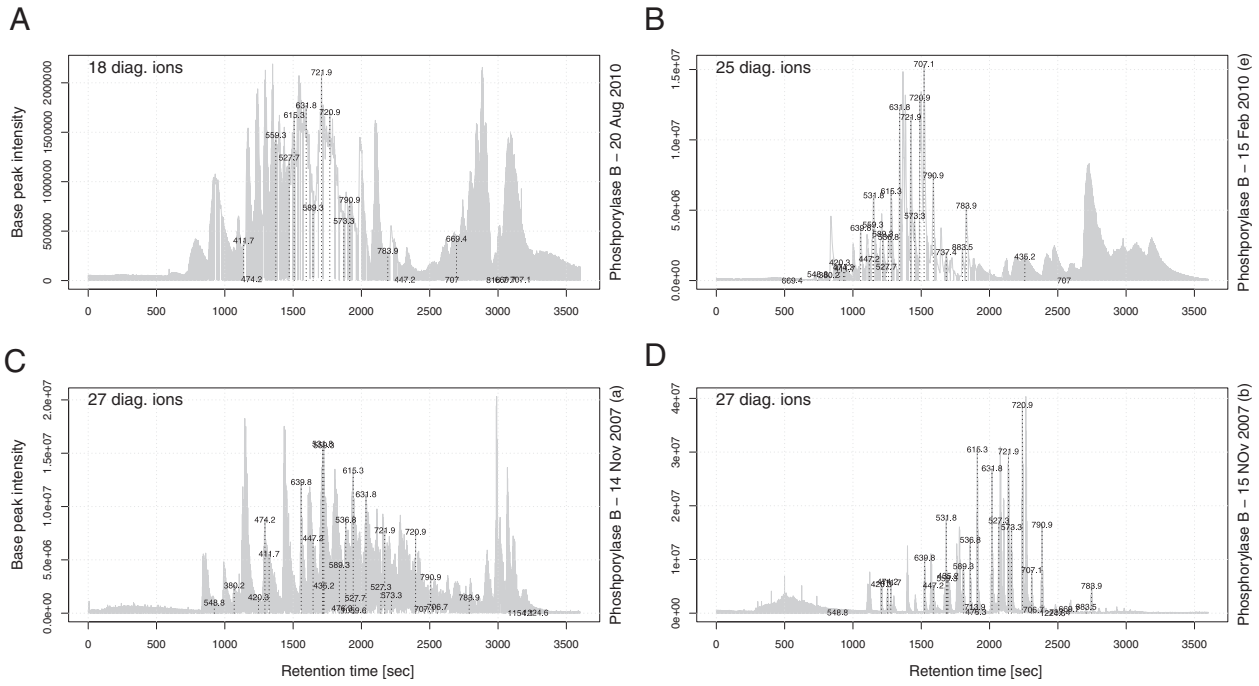

Figure 2. Four automatically generated chromatograms for the phosphorylase B standard indicating diagnostic ions as vertical dotted segments along with their MZ value. (A) This low quality chromatogram shows low base peak intensity and polyethylene glycol contamination, preventing any sample analysis. (B) This second example is of better quality but shows contamination of unknown origin starting after 2000 seconds that prompts for further inspection before proceeding with any sample. (C) This figure displays a chromatogram where, although many diagnostic ions have been identified, poor peak resolution is achived. (D) The forth example illustrates a good quality chromatogram, with good peak resolution and expected diagnostic ions retention times. These diagnostic plots do not replace the expert knowledge of mass spectrometrists, but represent a quick QC check and allow to easily trace a standard sample over time.

identified as well as their coverage as a first high-level assessment. More detailed analysis could monitor which peptides are observed and at what intensity. In terms of quantitative proteomics, facilities should be able to compare their measurement to a set of community-wide reference data, to statistically assess their results. Data from such standards can also be used for 'institutional memory', in other words, by having data on the same sample on the same instrument over time, a facility would be able to record the history of the performance of instruments, track batch numbers of reagents and operator error (Fig. 2). This would also represent an excellent test for newly trained staff. The continual running of standard samples comes at a price and funds are not always forthcoming for QC samples to be interrogated with necessary regularity. In addition, QC methods also need to be described to assess the quality of the data (Foster et al., submitted for publication).

Keeping up to date with new technologies can also be challenging. Reading journals is one mechanism, but possibly the best mechanism is attendance of workshops and specialist conferences. Yet again, funds are also not available to allow sufficient attendance of such events and certainly not available to all members of a core team.

Fundamental to a core facilities survival is not only being aware of new developments in terms of experimental advances, but also implementing these new advances into the day-to-day running of the facility. To implement new techniques into a facility requires a fairly large research effort to first test new methods and then to optimize them before they can be introduced into the routine workflow of the laboratory. It remains a challenge to try and balance the main core service work, the majority of which is fairly straightforward in terms of workflow and researching the more applied methods which could potentially be necessary for future core experiments.

Motivation and retention of core staff cannot be underestimated. Good proteomics requires skilled staff and a manager does not want to have to continually train new coworkers. Many aspects of the proteomics work flow are repetitive and unsatisfying. There is often a hazy delineation between what constitutes a collaborator or a client. Core facility members may not always get full recognition for their input. Exposure to a variety of different methodologies and their application, attendance at conferences, seeing through a project to the end and positive feedback from collaborators are all factors that may make the working life of a core team member more enjoyable.

Possibly, the biggest headache for core facilities, especially in the current economic climate, is funding. Some facilities will be centrally funded with an annual budget, others survive by making enough money from fee for service to cover costs, although rarely do these fees cover

new equipment and often the main priority for income into the facility is the salaries of staff, consumable costs and equipment maintenance. It is, therefore, becoming increasingly more common and necessary for core facilities to adopt a more business-like attitude to the day-to-day running of the lab. Key to this is the implementation of a business plan, which usually covers a three to 5-year period, and may be reviewed annually to ensure that targets are met for each year that has passed and if not, suitable alterations to the projected income or expenditure are made.

4 What would make life easier for core facilities?

The needs of newly forming facilities or in fact more established ones who struggle with some of the elements described above remain unmet on many levels. The following sections list our opinions on the key unmet needs of proteomics facilities.

4.1 Easy-to-use sample tracking and data storage

Many mechanisms exist, but most are sold commercially and are often difficult to justify to funding bodies [2]. There are some risks in terms of long-term maintenance with open source alternatives, especially when their user base is limited. However, buying software from a vendor does not guarantee for a better product or service. The requirements and utilization of a LIMS software can be shallow (in which case the investment might not be justifiable) or much more intensive, i.e. making use of many of the software's elaborate features. The latter may require the staff and users to adapt to new methods of recording data, resulting in some necessary constraints. After sufficient training and usage of the LIMS system, staff could become heavily dependent on the software for tracking experimental procedures. If problems arise from the normal functioning of such software, it can be time-consuming and expensive to rectify. Ideally, one dedicated person who is responsible for the LIMS and takes care of its configuration/updates etc. should be employed within the facility, resulting in additional costs.

UNMET NEEDS: The acceptance by funding bodies that it is as important to fund tracking and storage infrastructure as it is to fund consumables and large pieces of equipment.

4.2 Universal standards for each 'service' and community wide benchmarking studies

There is no accepted set of universal standards for proteomics. Part of the problem is that a proteomics facility will carry out many different types of application, each of which requires a well-characterized standard. For example, a facility may need a simple set of proteins or peptides to enable optimization of chromatography and instrument calibration and sensitivity. Another more complex standard may be needed to standardize and optimize more elaborate proteomics experiments. A well-documented phosphoproteomics standard, a membrane proteome standard, the list goes on. There is also undoubtedly a requirement for standards for relative and absolute quantitative proteomics types of experiments. Standards are not necessarily easy to produce. They must be manufactured in large quantities to overcome batch-to-batch variation, and be stable to allow them to still to be useful after varying storage conditions found in different facilities. They must also be very well characterized and the results of characterization will be available to all if they are going to have utility as community wide standards. Several such standards are available, for instance the Sigma protein mixes, Universal Proteomics Standard (UPS)1 and UPS2. The former developed in association with the Proteome Standards Research Group (sPRG) of the ABRF contains 48 human derived or recombinant human proteins each of which has been selected to limit heterogeneous post-translational modifications. This standard was further reformulated in UPS2 where the 48 proteins are available in six mixtures of eight proteins to present a dynamic range of five orders of magnitude, ranging from 50 pmol to 500 amol. A more complex mixture from Agilent has been created from a total cellular digest of *Pyrococcus furiosus*, which has approximately 2000 proteins. An issue with these standards is that as yet there is no mechanism to collate all the analyses of them collected by the community at large. UPS1 formed the subject of a benchmarking study by the ABRF in 2006, but the data collected from this study from over 74 labs were never made completely publically available and only an overview of the study was ever made available by the ABRF from their website (http://www.abrf.org/ResearchGroups/ProteomicsStandardsResearchGroup/).

Proteomics is woefully in need of a more thorough, well-characterized set of standards that are accessible to all and represents the wide gamut of applications associated with this technology. The commercially available standards may be prohibitively expensive for some facilities, but unfortunately the cost in making such a standard can only be supported by a commercial concern and perhaps the way forward is for funders of proteomics to realize the need for these standard and build the cost of their purchase and analysis into true cost of executing proteomics experiments.

Not only are standards needed to aid a facility's reproducible operation, another function is to enable inter facility testing and benchmarking. There have been some attempts at benchmarking studies within proteomics. The ABRF have been the driving force for many of these in the past. This organization has themed research groups made up of volunteers from the community and three proteomics-based research groups, Proteomics Research Group (PRG), Proteome Standards Research Group and Proteome Informatics Research Group (iPRG) who have organized bench-

marking studies in the past. These studies have been open to all and the presentation of the data collected has maintained the anonymity of the groups depositing data [3–6]. The studies have covered such areas as de novo sequencing and post-translation modification analysis, quantitation and informatics studies. A more recent independent study from Bell et al. involved a limited set of proteins sent for processing by 27 proteomics facilities worldwide including some of the best established. Of the 27 labs, members of only 7 labs initially reported all 20 proteins correctly. Thorough centralized analysis of the raw data showed that all proteins had been detected but missed identifications; environmental contamination, database matching and curation of protein identifications were the source of the initial failures [7]. The results of this study thus showed shortcomings even in the analysis from expert laboratories. The second part of the study involved requesting the participating laboratories to identify peptides of mass 1250 Da present in each protein; only one laboratory was successful although each laboratory provided raw data in which these ions could be found during the centralized analysis of the data. Each participating laboratory quickly improved the interpretation of their data with tuition from the centralized laboratory, indicating the value of taking part in such benchmarking studies (John Bergeron – personal communication).

Unfortunately, it was not a universal benchmarking study and therefore, although of interest, has little utility to the average core facility other than making them look at the efficiency of their own pipelines. The work that went into analysis of data from this study was immense and one could imagine an entire groups' worth of researchers would be needed to mine benchmarking data to the extent required for it to become a useful resource for the community. The National Cancer Institute (NCI) – Clinical Proteomic Technology Assessment for Cancer network (CPTAC) consortium has also recently published the results of an inter-facility study. They conducted a multi-laboratory study to assess reproducibility, recovery, linear dynamic range and limits of detection and quantification of multiplexed, MRM-based assays. Having created a standard sample and utilizing standardized protocols, they showed that these assays can be highly reproducible within and across laboratories and instrument platforms (to within 25% quantitative variation), and are sensitive to low microgram per milliliter protein concentrations in unfractionated plasma. [8]. This study not only allowed the use of a common set of samples for benchmarking, but it in fact benchmarked an entire analytic pipeline. This study involved only a subset of researchers all funded through the same NCI funded initiative and the samples generated cannot be publically requested.

A final study which could be thought of as a benchmarking study was carried under the auspices of Human Proteome Organization (HUPO) to test the reproducibility in 2-DE. The aim of the study is to make available reference protocols, images, image analysis and tools and samples for 2-DE. In the initial part of this study, samples made from a HeLa cell extract were distributed to 20 labs world-wide and the results from three gels per lab were compared across institutions. Both intra- and inter-lab reproducibility was demonstrated where a standardized approach was used. The current availability of this sample to the community as a whole is also unclear and there is no publication yet associated with this study other than poster abstracts.

Ideally, raw data files, summary results, meta-data and a standardized analysis pipeline (to make results easily comparable) for multiple assays and different platforms from universal standards should be centrally stored, documented and freely accessible. Optionally, anonymous data submission would occur on a voluntary basis to build a comprehensive and community-wide reference.

UNMET NEED: Availability of sets of universal standards covering all manner of sample types are currently not available to the community.

Perhaps, in future, HUPO will see through an initiative to provide standard sample sets, but this will require a great deal of resource and it is unlikely that any one organization will fulfil all the standards needs of the community.

4.3 Experimental designs and standardization of methods

All proteomics experiments start with researchers sitting down and planning the design of the experiment. A core facility may have little input into this. In larger scale experiments, replicates need to be employed to ensure that an observation is indicative of the biology and not just a technical artefact. In some cases, what should have been the most appropriate experimental design may only become apparent after the analysis of a set of data is complete. A lack of sufficient replicates or lack of sampling of the sets of proteins of interest may lead to data from which little can be interpreted. Researchers often enter into a different type of black box planning an experiment without access to the necessary information to ensure the most appropriate design as there are insufficient funds to carry out pilot experiments leading to data that will inform about a suitable design. Optimal experimental design requires input from many experts: biologists, specialists in the applied technology and statisticians. Funding for such sets of pilot experiments is frequently not forthcoming and researchers who apply for funding for experiments seem to be expected to have a sixth sense of the best design for their study. There is also a dearth of information about the experimental design in proteomics, with publications citing experimental designs without any discussion about why the design was utilized. How many papers' method's sections contain the following: 'three biological replicates for each condition were taken and the soluble proteins extracted'? Perhaps the missing words

here are 'three biological replicates for each condition were taken because that is all we had access to and the soluble proteins extracted although we had no idea whether the proteins likely to be of interest would be represented in this fraction'. Sometimes, it can be tough for a manager of a core facility to manage a situation where a researcher suggests experimentally flawed analysis, especially when the manager is under pressure to bring in revenue to his or her facility.

UNMET NEED: Funds for pilot experiments. More resources to aid efficient experimental design.

Working to a standard set of methods is not trivial. Many laboratories will have their own set of standard operating protocols and some procedures, for instance calibration of instruments is relatively easy to standardize; other procedures are not. Many current protocols commonly used in proteomics facilities still have room for further optimization and thus a standard protocol may be a moving target. Currently in the literature, for example, there are a bewildering set of protocols which aim to enrich the phosphoproteome. Each group will claim that their method is best in their hands, which is probably true, but this does not necessarily translate to other groups or be universal for all samples.

For facilities wanting to utilize a particular protocol for the first time, it would be highly desirable to have access to basic protocols even if these are then embellished and optimized by that facility with time. Currently, there is no centralized comprehensive source of such protocols and potential users are left sifting through the literature and individual laboratories web sites which may carry nuggets of information of how to get a particular procedure to work well. Resourcing a centralized frequently updated repository for common protocols may be challenging and outside the remit of most facilities. Perhaps, a proteomics-based organization such as HUPO or ABRF would be the most appropriate bodies to set up such a resource.

UNMET NEED: Centralized repository of protocols

4.4 Bioinformatics considerations

Just as the most effective laboratory-based protocol is a moving target, bioinformatics associated with proteomics is an even more dynamic entity.

4.4.1 Data analysis

Data analysis within proteomics comes in many flavours, from searching mass spectrometry data for the purpose of protein identification to extraction of quantitative information and statistical analysis. The former is fairly mature with numerous search engines available both from commercial sources and from open access. There is also a reasonable amount of information in the literature on the best way to utilize these searches and how to assess false discovery rate within identification of likely peptide sequences and post-translational modifications. There are also many different ways in which quantitative information can be extracted, with necessary software coming from both vendors and open source sites. Many of these softwares are black boxes with a lack of transparency of how they accomplish the spreadsheet of quantitative information which is frequently their output. There is a greater need for transparency for many of these softwares if they are to be used in the most effective manner, otherwise their misuse will go undetected leading to the literature potentially being polluted by false discoveries.

Statistical analysis of data is also poorly defined within the proteomics community. Data sets may need sophisticated and possibly customized analysis applied to them, but many researchers do not have access to the resources for this to be achieved and therefore may use inappropriate statistical analysis. For example, the application of well-known parametric tests like Student's *t*-test underlie important assumptions that must be met to obtain robust data: the data must be normally distributed, the different groups have to display homogeneous variances and the observations must be independent, and it is well known that several of these are not met in proteomics data sets. Sometimes, transformation of the data can leverage the requirements, but sometimes non-parametric tests have to be applied at the expense of the statistical power [9].

UNMET NEED: Better education tools and community-owned software for statistical approaches to proteomics data analysis.

4.4.2 Universal data formats, standardization of data reporting and data deposition

Multiple vendor or software specific formats can be difficult to cope with, especially when they are not fully described, easily parse-able or do not provide application programming interfaces (APIs) to query them.

Being at the source of data generation, core facilities deal with huge amounts of raw data and detailed information about the generation process for multiple technologies. They often use multiple vendor-specific software and formats and need to process the data and summarize the results in user-friendly formats for diverse assays and applications. Multiple formats and their associated software may result in long-term maintenance difficulties and in general, non-standard or closed formats require peculiar treatment before fitting into standard pipelines. Conversely, open standard formats are most efficiently incorporated into common pipelines and easily accommodated into new software. As such, the HUPO Proteomics Standards Initiative (PSI) initiative to develop standards that receive overall acceptance is welcome. One recent example is the new format mzML, for MS data [10]. Other examples are

mzIdentML for protein and peptide identifications, TraML for transitions in SRM and mzQuantML for quantitation data, all work in progress.

In terms of data reporting, the fine level of detail that can be extracted from raw data is generally not required by the user and often not desired. Decisions must thus be made in terms of QC, data processing, summarization and reporting for as many assays that are proposed. This in turn often requires multiple parallel software and pipelines. To reduce the maintenance load and streamline the analysis, which in turn increases the overall quality of the process, well-established standard formats are of great importance. Also, these are most welcome by software developers and bioinformaticians working in core facilities because of their wider acceptance, open description and allow the rationalization of the analysis pipeline.

Finally, core facilities may also be involved in data submission to public repositories. The latter work towards data standardization to ease their processing, annotation and mining activities and the fact that they allow easy submission of data in standardized form.

UNMET NEED: standardized formats for several data types, although under development, are not yet widely applied.

4.5 Educational tools

A comprehensive set of educational notes for proteomics has yet to be created. There are plenty of book and web sites that carry excellent descriptions of certain technologies [11–15] (www.fixingproteomics.org), but to date, there is not a 'one stop shop' where proteomics protocols are collated. Funding to maintain such a resource could be a challenge to secure.

In the above sections, it is clear that proteomics facilities would benefit from easy access to educational tools. These tools may take on the form of tutorials, (e.g. statistical test, experimental design), protocols and also bulletin boards to allow questions to be posed of the community. One of the most mature electronic discussion groups has been maintained for a good many years by the ABRF. There are many others, but their universal usage remains to be seen, with several of them having received no postings for several months/years.

Specialist methods forum type meetings are also of great value to core facility members who may not get much to take home and try out from plenary lectures given by the more prominent members of the proteomics community. Many core facilities do not have unlimited resources to send their staff to meetings, thus such forum meetings should be local, cheap and concentrate on the 'how to do' aspects of proteomics. In the UK, the Proteomics Methods Forum is an annual meeting which was established in 2005. It is free to attend, usually held in an academic location and the speakers taken from the laboratory members who actually carry out protocols on a daily basis.

UNMET NEED: Specialist conferences and easily accessible centralized educational tools

5 What the future may hold for proteomics core facilities

Core facilities should be, without doubt, considered very valuable resources within academic and industrial research environments. A centralized core of expertise means that non-experts who wish to incorporate proteomics into their research may approach the facility for expert advice in terms of discussing the merits/flaws of proteomics for their research, planning experiments, budgeting costs, possible outcomes and technical details. This is an invaluable resource to a proteomics novice who is keen to gain knowledge rapidly without having to resort to analysing numerous publications (which may or may not be relevant) to understand basic principles. Staff within such facilities often have a broad range of skills including expertise in sample preparation, analytical chemistry, method development, bioinformatics and statistics. A potential novice user of the facility will therefore have a range of expertise at their disposal, and crucially, within the same location. In addition, due to the fact that full proteomic experiments usually consist of different experiments/analyses performed by the aforementioned experts who work under the same roof, the transfer between experiments/analyses is relatively seamless and hence timesaving and cost-effective. A further advantage is that most facilities have tried and tested protocols which have been in place within the workflow for significant amounts of time. As a consequence of this, data should be extremely reproducible and consistent across a number of different types of experiment.

With a down-turn in the global economy, and uncertainty about the future of funding of science, proteomics core facilities are likely to experience a tough few years. Many are already run as small businesses in an academic setting particularly and this model may become more prevalent amongst core facilities.

There has been much discussion in the literature about what proteomics has achieved to date. Perhaps, it was oversold in its infancy, particularly with respect to the speed with which this technology would reveal useful biomarkers of disease. A recent article, however, suggests that proteomics is now 'ready for the big time' and provided that technologies are applied carefully and correctly, ensuing data can be highly informative and reproducible [16].

It is clear that for proteomics to move forward efficiently particularly in the setting of a core facility, there are several unmet needs which need to be addressed to assist in achieving streamlined, high-quality proteomics services. These largely centre around standardization of

protocols, data handling and analysis, educational tools, across community standard samples and benchmarking studies.

We finally challenge vendors and more established facilities to democratize proteomics such that high standards are achievable by all in a facile and well-documented manner to promote proteomics and improve the current perception of it as the underachieving 'omic technology'.

The authors thank the following people for their comments during the preparation of this manuscript. The authors thank John Bergeron, Chris Turck for helpful discussions while preparing this manuscript. L. G. is a BBSRC funded post-doctoral research fellow (BB/G024618/1).

The authors have declared no conflict of interest.

6 References

[1] Wilkins, M. R., Arthur, J. W., Junius, F. K., Harrison, M. J. et al., Information management for proteomics: a perspective. *Expert Rev. Proteomics* 2008, *5*, 663–678.

[2] Piggee, C., LIMS and the art of MS proteomics. *Anal. Chem.* 2008, *80*, 4801–4806.

[3] Arnott, D., Gawinowicz, M. A., Grant, R. A., Neubert, T. A. et al., ABRF-PRG03: phosphorylation site determination. *J. Biomol. Tech.* 2003, *14*, 205–215.

[4] Arnott, D., Gawinowicz, M. A., Kowalak, J. A., Lane, W. S. et al., ABRF-PRG04: differentiation of protein isoforms. *J. Biomol. Tech.* 2007, *18*, 124–134.

[5] Falick, A. M., Kowalak, J. A., Lane, W. S., Phinney, B. S. et al., ABRF-PRG05: de novo peptide sequence determination. *J. Biomol. Tech.* 2008, *19*, 251–256.

[6] Turck, C. W., Falick, A. M., Kowalak, J. A., Lane, W. S. et al., The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 Study. *Mol. Cell. Proteomics* 2007, *6*, 1291–1298.

[7] Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E. et al., A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Meth.* 2009, *6*, 423–430.

[8] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J. et al., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotech.* 2009, *27*, 633–641.

[9] Karp, N. A., Lilley, K. S., Design and analysis issues in quantitative proteomics studies. *Proteomics* 2007, *7*, 42–50.

[10] Martens, L., Chambers, M., Sturm, M., Kessner, D. et al., mzML – a community standard for mass spectrometry data. *Mol. Cell. Proteomics* 2010.

[11] Peirce, M. J., Wait, R. (Eds.), *Membrane Proteomics*, *Vol. 528*, Springer, New York 2009, ISBN: 978-1-60327-309-1.

[12] Reinders, J., Sickmann, A. (Eds.), *Proteomics: Methods and Protocols*, *Vol. 564*, Springer, 2009, ISBN: 978-1-60761-156-1.

[13] Thiellement, H., Zivy, M., Damerval, C., Méchin, V. (Eds.), *Plant Proteomics: Methods and Protocols*, *Vol. 355*, Springer, 2006, ISBN: 978-1-58829-635-1.

[14] Thompson, J. D., Schaeffer-Reiss, C., Ueffing, M. (Eds.), *Functional Proteomics: Methods and Protocols*, Springer, *Vol. 484*, 2008, ISBN: 978-1-58829-971-0.

[15] Vlahou, A. (Ed.), *Clinical Proteomics: Methods and Protocols*, *Vol. 428*, Springer, 2008, ISBN: 978-1-58829-837-9.

[16] Nilsson, T., Mann, M., Aebersold, R., Yates, J. R. et al., Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Meth.* 2010, *7*, 681–685.

