

alternative chromophores with extended conjugation in order to access the near-infrared. Perhaps the solution is to abandon the fluorescent protein superfamily entirely and look to alternative protein scaffolds for development of near-infrared fluorescent proteins. Such an approach was recently taken by Tsien and co-workers in the development of an infrared fluorescent protein based on a bacteriophytochrome protein that binds an ubiquitous biliverdin chromophore¹⁹.

The rise of fluorescent proteins as powerful imaging tools has been marked by fairly regular surprising revelations regarding the versatility and utility of these remarkable fluorescent probes. What other surprises do fluorescent proteins have in store? Although we cannot answer that question with any degree of certainty, one thing is clear: when fluorescent proteins next reveal an interesting new property, the research community will once again seize it and exploit it to enable exciting new applications on the cutting edge of biological imaging technology.

COMPETING INTEREST STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturemethods/>.

1. Tsien, R.Y. *Annu. Rev. Biochem.* **67**, 509–544

- (1998).
2. Matz, M.V. *et al. Nat. Biotechnol.* **17**, 969–973 (1999).
 3. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. *Nat. Methods* **2**, 905–909 (2005).
 4. Wang, L., Jackson, W.C., Steinbach, P.A. & Tsien, R.Y. *Proc. Natl. Acad. Sci. USA* **101**, 16745–16749 (2004).
 5. Merzlyak, E.M. *et al. Nat. Methods* **4**, 555–557 (2007).
 6. Shaner, N.C. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
 7. Shaner, N.C. *et al. Nat. Methods* **5**, 545–551 (2008).
 8. Kredel, S. *et al. PLoS One* **4**, e4391 (2009).
 9. Shcherbo, D. *et al. Nat. Methods* **4**, 741–746 (2007).
 10. Shcherbo, D. *et al. Biochem. J.* **418**, 567–574 (2009).
 11. Strack, R.L. *et al. Nat. Methods* **5**, 955–957 (2008).
 12. Ai, H., Hazelwood, K.L., Davidson, M.W. & Campbell, R.E. *Nat. Methods* **5**, 401–403 (2008).
 13. Tomosugi, W. *et al. Nat. Methods* **6**, 351–353 (2009).
 14. Patterson, G.H. & Lippincott-Schwartz, J. *Science* **297**, 1873–1877 (2002).
 15. Subach, F.V. *et al. Nat. Methods* **6**, 153–159 (2009).
 16. McKinney, S.A., Murphy, C.S., Hazelwood, K.L., Davidson, M.W. & Looger, L.L. *Nat. Methods* **6**, 131–133 (2009).
 17. Valentin, G. *et al. Nat. Methods* **2**, 801 (2005).
 18. Kremers, G.J., Hazelwood, K.L., Murphy, C.S., Davidson, M.W. & Piston, D.W. *Nat. Methods* **6**, 355–358 (2009).
 19. Shu, X. *et al. Science* **324**, 804–807 (2009).

technological improvements in proteomics will go a long way to overcome these difficulties¹, but they need to be accompanied by rigorous analysis. Three ‘analysis’ papers published in the last five years in *Nature Methods* have broken new ground in investigating crucial data quality issues in the proteomics field. One of them deals with protein identification, another with the enrichment of phosphorylated peptides, and the third with the evaluation of proteomics researchers themselves!

A main developmental direction of our discipline is pushing the identification of ever more proteins in complex proteomes, something to which MS is uniquely suited. However, many of the early landmark papers in the last 5–10 years that established the feasibility of large-scale protein identification were obtained on low-resolution instruments and without proper statistical analysis. We now know that a large proportion of the identifications obtained from such projects were in fact false positives. For example, peptide lists contained a large proportion of nontryptic peptides, whereas it is now generally acknowledged that trypsin, at least in proteomics experiments, is extraordinarily sequence-specific². The recognition of these data quality issues prompted a gradual, though still not complete, switch to high-resolution techniques. It also lent impetus to efforts to standardize the reporting of proteomics protocols and data, and to the development of bioinformatics tools to directly determine the false positive rate independently of the peptide database search score.

Aebersold and co-workers addressed this issue early on by developing an algorithm that decomposed scored distributions into underlying false positive and true positive distributions and assigned likelihoods that peptides with a given database identification score were in fact correctly identified. This development, implemented in their PeptideProphet and ProteinProphet software^{3,4}, was an important step in bringing some rigor to the identification process in low-resolution data. Even more simple and powerful—and applicable to high- as well as low-resolution data—was the concept of applying reverse sequence databases to determine false positive identification rates. This approach is very straightforward and only involves searching the data against the normal or ‘forward’ database and against the sequence-reversed database (also called a ‘target-decoy’ database, if it is

Comparative analysis to guide quality improvements in proteomics

Matthias Mann

The potential of mass spectrometry-based proteomics to advance biology and biomedicine is nearly unlimited but so is its potential for generating bad data. Apart from the pursuit of technological progress in protocols and instruments, stringent comparative analyses of different approaches are critical for fully developing the discipline.

Systems-wide analysis of any biological entity is hard. For proteins it is particularly daunting because proteomics lacks an equivalent to hybridization assays based on Watson-Crick base pairing or amplification reactions based on PCR. Fortunately, we do have the wonderful tool of mass spectrometry (MS) at our

disposal. MS is one of the most versatile technologies used in biology, and continuing radical improvements in the technology amaze even the most seasoned observers. However, not all is well in the discipline of proteomics, and much fuzzy thinking and bad data have unfortunately found their way into the literature. Purely

Matthias Mann is at the Max Planck Institute for Biochemistry, Martinsried, Germany.
e-mail: mmann@biochem.mpg.de

concatenated with the forward database⁵). Once 1% of reverse hits have accumulated, the acceptable threshold score for 99% accuracy is reached, regardless of statistical distribution of errors, workflow and mass spectrometer used, database search score and so on. Although mathematically more sophisticated approaches may in principle be more efficient⁶, the simplicity and robustness of reverse sequence searching make it easily accessible to members of all proteomics laboratories.

The Gygi group has played an important role in developing the reverse database approach⁵, and their analysis paper in *Nature Methods* in 2005 nicely illustrates the power of this approach for comparing different proteomics workflows⁷. In this work, Gygi and colleagues asked three different questions: is a linear ion trap or a quadrupole time of flight instrument better at identifying peptides? Is the Mascot⁸ or the Sequest⁹ database search algorithm superior? And finally, how useful is it to rerun samples to increase peptide identification? Such questions are notoriously difficult to answer in a general and defensible way because results depend on many parameters. However, using the reversed-sequence database strategy to keep the false positive rate fixed at 1%, Gygi and co-workers determined that the two instruments yielded data of comparable quality, but the linear ion trap was much faster. Mascot and Sequest were also roughly similar in performance, but Mascot was better-suited to higher-resolution data, and Sequest was better-suited to low-resolution and noisy data. Peptides identified in repeat measurements contained exceedingly few false positives; false positives were likely to be those peptides only identified in a few runs. This last finding means that identifications in large-scale projects with many repeats and redundancy cannot just be added together on a run-by-run basis; instead, all the data need to be analyzed together to guarantee a uniform false positive rate.

The publication of this paper coincided with the end of the reign of low-resolution ion traps and the quadrupole time of flight instruments in proteomics and just preceded the reign of the hybrid ion trap-Fourier transform instruments, which still continues today. These instruments add high resolution—indispensable for accurate quantification—and very high mass accuracy to the capabilities of the linear ion trap. With these powerful instruments, the majority of



Figure 1 | Is it the car (the mass spectrometer) or the driver (the proteomics researcher) that wins the race (characterizes the proteome correctly)? At least for simple protein mixtures it is the driver and not the car, as analyses from HUPO and ABRF demonstrate.

fragmented peptides can now be identified using workflows and software that make full use of the new instrumental capabilities¹⁰.

Another major application to which mass spectrometry is uniquely suited is the large-scale and accurate analysis of post-translational modifications (PTMs). Low site stoichiometry of PTMs and their dilution in the general pool of peptides necessitate specific enrichment of modified peptides. For phosphorylation, the most commonly studied modification, there are many different protocols for enrichment, all of which exploit the distinct nature of the phospho-group as a 'handle'. For reasons that are not entirely clear, these protocols achieve very different depth of coverage in different laboratories, and, as a result, each group tends to be most successful with their favorite protocol. In a 2007 paper in *Nature Methods*, Aebersold and co-workers systematically evaluated three of the most popular phosphoprotein enrichment protocols¹¹. They subjected a cytosolic extract of a *Drosophila melanogaster* cell line to enrichment by phosphoramidate chemistry (PAC), immobilized metal affinity capture (IMAC) or titanium dioxide supplemented with 2,5-dihydroxybenzoic acid (DHB-TiO₂) or phthalic acid, and measured each fraction six times with liquid chromatography-tandem MS (LC-MS/MS). The researchers identified less than 1,000 phosphorylation sites, a relatively small fraction of the phosphoproteome, and found that each method contributed distinct phos-

phopeptides to the total. Thus, although technical replicates were quite reproducible, the different enrichment protocols did not yield highly overlapping results. Each enrichment method appeared to isolate different populations of phosphopeptides, although a common theme for these differences was not apparent. Aebersold and colleagues thus concluded that no single phospho-enrichment strategy can cover the entire phosphoproteome.

Though this work is a good example of the type of analysis that needs to be performed to establish optimal workflows, the results illustrate some of the difficulties in comparing complex proteomics workflows. For example, the authors determined only 156 phosphorylation sites from six runs using the DHB-TiO₂ methods, whereas they found three times as many sites with the IMAC method. In contrast, other groups have found DHB-TiO₂ to be superior to IMAC. Our group, for example, routinely uses DHB-TiO₂ to identify thousands of phosphorylation sites¹². In a subsequent publication, Aebersold and colleagues reported identification of tenfold more phosphopeptides from *Drosophila* using the same combination of the three enrichment protocols¹³. However, a study using the DHB-TiO₂ method alone resulted in the same number of *Drosophila* phosphorylation sites¹⁴. So the jury is still out on whether separate isolation protocols are the strategy of the future or if deeper sampling with one optimized method will go just as

far. It would be very interesting to repeat this study using quantitative proteomics approaches: one could use SILAC to label *Drosophila* cells¹⁵, enrich the phosphopeptides by the three different protocols and then directly judge the quantitative differences in phosphopeptide abundance. I would expect that the large majority of phosphopeptides would indeed be present in all three methods, just at different abundances.

The above two examples deal with challenging technical problems in proteomics—one of which is well on its way to being solved and one of which is still in flux. The 2009 analysis by Bergeron and colleagues in *Nature Methods* deals with a more sociological issue: ignoring for a moment the cutting-edge publications in high-impact journals, what is the actual state of the art among practicing proteomics laboratories? To answer this question they sent a test sample to 27 different groups and asked them to identify the constituent proteins¹⁶. Given that entire proteomes can now be comprehensively identified and quantified¹⁷, the protein test sample seemed to be absurdly simple: it contained only 20 proteins, all at the same molar concentration. However, this made the results all the more shocking: only members of a few of the laboratories identified all 20 proteins. Several groups missed a number of proteins or reported proteins that were not actually present in the test sample. Bergeron and co-workers charitably attributed these problems mainly to database issues. Indeed, with some ‘hand-holding’ members of the participating laboratories all identified the 20 proteins in a second round, and a centralized re-analysis of the data revealed that most of the groups had the data in hand but failed in their analysis of it. But this unexpected result shows just how large a gradient exists between the few expert laboratories that publish in high-impact journals and the many laboratories that have to make do with limited resources and expertise. One might expect these results to be caused by outdated equipment but this was not so: experienced researchers with older mass spectrometers

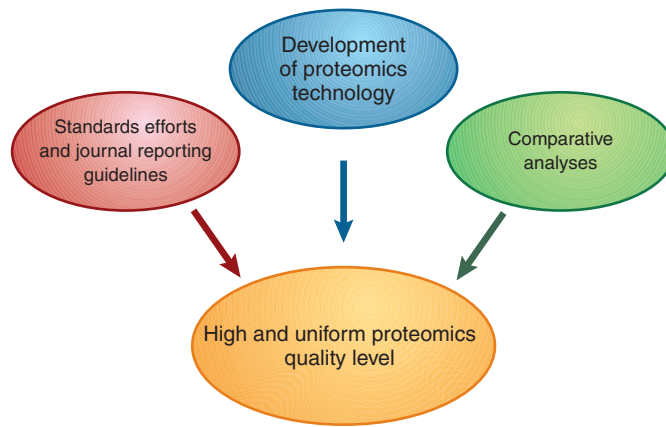


Figure 2 | Technological developments have to be accompanied by reporting standards and comparative analyses to improve data quality in proteomics.

performed better than less-experienced researchers with the latest equipment (Fig. 1). Neither are these results likely to be flukes because a similar study by the American Biomolecular Resources Association (ABRF) came to the same general conclusions (ABRF standards group, <http://www.abrf.org/ResearchGroups/ProteomicsStandardsResearchGroup/EPosters/sPRGStudy2006OralPresentation.pdf>; 2006). Clearly, the community has its work cut out to bring the majority of proteomics laboratories to the same standard now achieved by the most advanced ones.

Thoughtful and well-performed analysis studies serve a very important function in a technology-driven discipline that is applied by a diverse set of researchers. Such studies are difficult to perform such that the findings are sound and generally applicable, but they provide much-needed guidance in a fast-changing field. They synergize with technological advances and the development of community standards to raise quality standards in proteomics (Fig. 2).

In the future, several areas of proteomics apart from the ones discussed here would be well served by such studies. This includes quantitative proteomics, for which there is a continuing need to benchmark different methods and workflows against each other; targeted proteomics, for which there is a need to evaluate new techniques such as multiple reaction monitoring¹⁸, particularly the false positive rates of the technique, not to mention the huge and increasingly important theme of PTM analysis, which

includes relative quantification and the determination of the stoichiometry of modification. Furthermore, as proteomics increasingly becomes capable of analyzing *in vivo* samples, the issue of handling and analyzing very small tissue amounts will come to the fore. Last but not least, the increasingly sophisticated and demanding downstream bioinformatic analysis of the proteomic data will increasingly need quality standards¹⁹. All these developments will take place while the proteomic user base will continue to grow

exponentially and increasingly include the commercial and medical communities in addition to the academic research community. If the history of the last five years is any guide, *Nature Methods* and other journals will have a crucial role during the next five years in helping to develop and safeguard quality standards for proteomics.

1. Mann, M. & Kelleher, N.L. *Proc. Natl. Acad. Sci. USA* **105**, 18132–18138 (2008).
2. Olsen, J.V., Ong, S.E. & Mann, M. *Mol. Cell. Proteomics* **3**, 608–614 (2004).
3. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. *Anal. Chem.* **74**, 5383–5392 (2002).
4. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. *Anal. Chem.* **75**, 4646–4658 (2003).
5. Elias, J.E. & Gygi, S.P. *Nat. Methods* **4**, 207–217 (2007).
6. Kim, S., Gupta, N. & Pevzner, P.A. *J. Proteome Res.* **7**, 3354–3363 (2008).
7. Elias, J.E., Haas, W., Faherty, B.K. & Gygi, S.P. *Nat. Methods* **2**, 667–675 (2005).
8. Perkins, D.N., Pappin, D.J.C., Creasy, D.M. & Cottrell, J.S. *Electrophoresis* **20**, 3551–3567 (1999).
9. Eng, J.K., McCormack, A.L. & Yates, J.R. III *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
10. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
11. Bodenmiller, B., Mueller, L.N., Mueller, M., Domon, B. & Aebersold, R. *Nat. Methods* **4**, 231–237 (2007).
12. Olsen, J.V. *et al. Cell* **127**, 635–648 (2006).
13. Bodenmiller, B. *et al. Mol. Syst. Biol.* **3**, 139 (2007).
14. Hilger, M., Bonaldi, T., Gnäd, F. & Mann, M. *Mol. Cell. Proteomics* **8**, 1908–1920 (2009).
15. Bonaldi, T. *et al. Mol. Cell* **31**, 762–772 (2008).
16. Bell, A.W. *et al. Nat. Methods* **6**, 423–430 (2009).
17. de Godoy, L.M. *et al. Nature* **455**, 1251–1254 (2008).
18. Addona, T.A. *et al. Nat. Biotechnol.* **27**, 633–641 (2009).
19. Kumar, C. & Mann, M. *FEBS Lett.* **583**, 1703–1712 (2009).