Cross-linking/mass spectrometry as a new field and the proteomics information mountain of tomorrow

EXPERT REVIEWS

Cross-linking/mass spectrometry as a new field and the proteomics information mountain of tomorrow

Expert Rev. Proteomics 9(5), 485–487 (2012)

Juri Rappsilber

Wellcome Trust Centre for Cell Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK
and
Department of Biotechnology, Technische Universität Berlin, 13353 Berlin, Germany
juri.rappsilber@ed.ac.uk;
juri.rappsilber@tu-berlin.de

The European Proteomics Association (EuPA) 2012 Scientific Congress 'New Horizons and Applications for Proteomics', hosted by the British Society for Proteome Research (BSPR)
Glasgow, Scotland, UK, 12 July 2012

Cross-linking/mass spectrometry ended decades of method developments and entered the era of applications at this year's European Proteomics Association meeting. The train has started moving, with successful applications of this tool by multiple pioneering laboratories addressing biological and structural problems. Proteomics, on the other side, sees ever increasing data volumes, leading to questions as to how to store the data mountain publically, use it and convert it into testable hypotheses. The European Proteomics Association meeting has been complementary to the American Society for Mass Spectrometry meeting in many ways, also thanks to its more manageable size and the vision of the organizers in inviting some of Europe's best emerging minds.

Keywords: cross-linking • data mountain • proteomics • science policy

To me, this has been one of the most inspiring meetings for a long time. Regularly attending the ASMS meeting is an essential experience in proteomics. Yet even after years of doing so it is difficult not to be stunned and simultaneously be paralyzed by its now 6277 participants, 2998 posters and up to eight parallel sessions (ASMS 2012). The European Proteomics Association (EuPA) conference 2012 was more tailored to networking. With 393 participants, 262 posters and only one session at a time one could actually meet people multiple times. The organizers had attracted some of the best emerging scientists in the field that Europe can muster. Quite possibly, this is the best meeting for PhD students to hunt for an exciting proteomics postdoc position. I personally did not miss a dominance of political heavy weights and scientific super heroes. Adjectives to my liking are cool and hip as well as raw, hungry and searching.

Cross-linking/mass spectrometry as an emerging field

My personal highlight: The Workshop 'Cross-linking/mass spectrometry' with James Bruce (University of Washington, WA, USA), Florian Stengel (Aebersold laboratory, ETH Zurich, Switzerland) and Zhuo Angel Chen (Rappsilber laboratory, University of Edinburgh, UK, and TU Berlin, Germany) possibly marked the endpoint of four decades of frustration. Cross-linking may be one of the oldest biochemical tools for the investigation of protein interactions and structure [1]. It certainly has not been one of the most successful ones for long. Even if you have tried it few years ago, likely you have been disappointed by elusive cross-links that did not yield to mass spectrometric detection. Early departures from this, certainly in wealth of data, include analyses of multiprotein complexes such as Nup84c (four subunits, 180 kDa [2]), Pol II (12 subunits, 500 kDa [3]) and Pol II-TFIIF

 10.1586/EPR.12.44 ISSN 1478-9450

(15 subunits, 670 kDa [3]). Moving on, data on the proteasome (33 different prot-eins, 47 subunits, 2.5 MDa) helped to model high-resolution structures into a cryoEM map [4]. In a handful of laboratories the tool has matured in parallel, to a level that now allows routine use in structural and mechanistic research.

Florian Stengel showed in Glasgow how cross-link data could even correct the incorrect subunit arrangement of a crystal structure [5]. This somewhat surprising initial arrangement in the TRiC/CCT complex was the result of the high similarity of the subunits involved and low resolution of the original EM densities and crystal diffractions. Ultimately, integrated structural biology is the way forward to study multiprotein complexes, using any data available [6]. Making use of cross-link constraints, such as providing peptide-level resolution, thus fills the gap between high and low resolution methods [1].

James Bruce, a pioneer of complex mixture analysis by cross-linking/mass spectrometry [7–9], stretched the vision of cross-linking/mass spectrometry to *in vivo* analyses of *Escherichia coli* and human cells. He reported the impressive number of over 1000 observed cross-links in either. Of course, the tip of the interactome iceberg may still hardly be scratched. Still, add the nearly 7000 cross-links reported by the Rappsilber lab in the 'Chemical Proteomics' session for the less complex *Mycoplasma pneumoniae* lysate and it may become clear, nevertheless, that even smaller but still complex heterogeneous structures and networks are starting to fall within reach of this tool. The challenge now will be for the pioneers to obtain ever more data in whole cell lysates or from intact cells. However, the tools are solidly established already and wait for adaptation by other laboratories for more routine, but possibly also biologically more exiting mechanistic studies.

In theme with the workshop, Ruedi Aebersold (ETH Zurich, Switzerland), a member of the League of Extraordinary Gentlemen in proteomics, presented such a biological showcase of cross-linking/mass spectrometry during his award talk as the EuPA Senior Scientist 2012. He gave a thrilling outlook into what may happen soon in many molecular biology laboratories around the globe when combining cross-linking/mass spectrometry with bead-based isolation of protein complexes. This is also backed by independent work from the Rappsilber laboratory by Zhuo Angel Chen [10].

Adding quantitation, Zhuo Angel Chen presented in the workshop the use of isotope labeled cross-linkers to study conformation changes of proteins. She focused on the human complement system as a conformation-controlled network and proposed the domain arrangement of complement C3(H_2O) based on cross-link data and available crystal structures of two conformers with different PTM status, C3 and C3b. Exciting times are ahead and the number of possible applications of cross-linking/mass spectrometry are countless. I foresee a great future for cross-linking/mass spectrometry as an emerging field and I look forward to a culture in which key developments and inspirations by others are acknowledged and embraced.

Transforming data mountains into hypotheses

Rune Linding (Technical University of Denmark) presented an awe-inspiring conversion of proteomic data mountains into information mountains, requiring multiple dedicated experimental and computational tools, for example, NetworKIN [11] besides mass spectrometers and expertise from informatics to medicine. The emerging field of network medicine is taking off and proteomics is playing a central role in that [12]. While such studies may require multidisciplinary specialized environments they are archetypical to what is happening to large parts of biological science in order to address some of the big questions. An interesting discussion followed, seeing biology as the new physics by relying on large teams of scientists that, indeed, may need a different publication model. Pubmed knows only one first and one last author regardless of how many stars are given for equal contribution and shared communication. Is big science in biology driving students and postdocs out of their careers as only one can be first author? Maybe there is more to be learned from physics, besides teamwork leading up to a publication.

Regardless of its impact, the fate of mass spectrometric data is currently all too often inaccessibility and being lost. A team of scientists spearheaded by the EBI is working on PRIDE, a proteomics data repository, to address this [13]. While uploads may have been cumbersome in the past, new data analysis tools allow direct interaction with PRIDE. For example, Lennart Martens' group (Ghent University, Belgium) presented integrative tools with an extensive user interface that allow using multiple search tools (OMSSA, XTandem, Mascot) in one analysis and presenting the combined results to the user in an interactive analysis environment with one-click submission to PRIDE. Check out SearchGUI [14] and PeptideShaker [101]. Angus Lamond's laboratory (University of Dundee, UK) is going one step further and has created a repository that stores all of their experimental data together with detailed metadata annotations for integrative exploration. Angus announced that this 'PepTracker' software [102] is currently being tested at the University of Dundee and will soon be available for general distribution. This creates a basis for 'super experiments', where the integration of richly annotated data and results from many separate experiments reveals new relationships between proteins and cellular mechanisms and can generate hypotheses that were not foreseen in the design of any of the original experiments [15,16]. Meta-analyses, such as those used to define functionally interesting proteins in mitotic chromosomes [17], often rely on machine learning. As the analysis challenge of identifying and quantifying proteins is slowly being addressed, the focus of bioinformaticians will shift increasingly towards data integrating approaches for hypothesis generation. This will likely shift job descriptions in proteomics from mass spectrometry experience to data handling and hypothesis generation skills. Students are well advised to prepare through adequate choice of courses and more computational researchers will (hopefully) make their way into biology.

An integrated EuPA vision

The EuPA congress organizers focused on emerging science and scientists, giving this meeting a fresh and unique feel. EuPA could consider picking up on this vision by also awarding an 'Emerging Scientist' prize. I argue here that this would have a larger impact on European proteomics than any other prize, senior or junior. The possibly most challenging stage in a scientist's career is the

consolidation phase; junior funding sees its ends and senior funding is only a hope on the horizon. This is the point when a scientist must show vision yet sees it confronted with the often too conservative thinking of journals and/or funding agencies. The field could strengthen the backing of mid-career scientists by a solid slap on the shoulder and a 'carry on, we see that you are heading in the right direction', just before mid-career turns into fame. For EuPA, in awards and its congress, a key aim might become to scout for and foster budding and emerging talent.

To poster authors

A comment on the more technical side, poster authors may want to have a look at some of the company posters. Some of the clearest posters came from companies. I would advise students to pair up and measure the time it takes one of them to extract a message from a poster that they can explain to the other. If it takes more than 10 s for a first idea, the poster is poorly made. It took me about that time to see that poster 1 [18] had niftily exploited high MS2 resolution and the mass difference between ^{15}N and ^{13}C to increase the multiplexing of TMT after two of the TMT reagents saw a change in their labeling scheme most recently. My prize for the most original poster design would go to the laboratory of Concha Gil [103]. Even if it took me much longer than 10 s to extract content from their posters, they have my vote for being refreshingly different. See for yourself when you get a chance.

As workshop organizer, speaker, session head and poster prize judge I possibly came closest to being an organizer without actually having been one. I was given these pages to give a personal account of this year's EuPA meeting. I apologize for my very subjective take on the meeting and to all those colleagues whose work I did not mention in this report. The French Electrophoresis and Proteomics Society [104] will host the 7th EuPA Congress in Saint Malo, France, in 2013.

Financial & competing interests disclosure

J Rappsilber thanks all researchers named in this report for comments and Hanno Steen for critically reading the manuscript. J Rappsilber is generously supported by the Wellcome Trust through a Senior Research Fellowship (084229), a Wellcome Trust Centre Core Grant (092076) and an instrument grant (091020), and the Einstein Foundation. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Rappsilber J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* 173(3), 530–540 (2011).

2 Maiolica A, Cittaro D, Borsotti D *et al.* Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell Proteomics* 6(12), 2200–2211 (2007).

3 Chen ZA, Jawhari A, Fischer L *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* 29(4), 717–726 (2010).

4 Lasker K, Förster F, Bohn S *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl Acad. Sci. USA* 109(5), 1380–1387 (2012).

5 Leitner A, Joachimiak LA, Bracher A *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* 20(5), 814–825 (2012).

6 Alber F, Förster F, Korkin D, Topf M, Sali A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* 77, 443–477 (2008).

7 Tang X, Munske GR, Siems WF, Bruce JE. Mass spectrometry identifiable cross-linking strategy for studying protein–protein interactions. *Anal. Chem.* 77(1), 311–318 (2005).

8 Zheng C, Yang L, Hoopmann MR *et al.* Cross-linking measurements of *in vivo* protein complex topologies. *Mol. Cell Proteomics* 10(10), M110.006841 (2011).

9 Bruce JE. *In vivo* protein complex topologies: sights through a cross-linking lens. *Proteomics* 12(10), 1565–1575 (2012).

10 Chen ZA, Rasmussen M, Tahir S, van der Sar S, Hardwick KG, Rappsilber J. Cross-linking analysis of affinity-purified multi-protein complexes. Presented at: *56th ASMS Conference on Mass Spectrometry*. Denver, CO, USA, 1–5 June 2008.

11 Linding R, Jensen LJ, Ostheimer GJ *et al.* Systematic discovery of *in vivo* phosphorylation networks. *Cell* 129(7), 1415–1426 (2007).

12 Erler JT, Linding R. Network medicine strikes a blow against breast cancer. *Cell* 149(4), 731–733 (2012).

13 Wang R, Fabregat A, Ríos D *et al.* PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat. Biotechnol.* 30(2), 135–137 (2012).

14 Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* 11(5), 996–999 (2011).

15 Lamond AI, Uhlen M, Horning S *et al.* Advancing cell biology through proteomics in space and time (PROSPECTS). *Mol. Cell Proteomics* 11(3), O112.017731 (2012).

16 Boulon S, Ahmad Y, Trinkle-Mulcahy L *et al.* Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. *Mol. Cell Proteomics* 9(5), 861–879 (2010).

17 Ohta S, Bukowski-Wills JC, Sanchez-Pulido L *et al.* The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* 142(5), 810–821 (2010).

18 Werner T, Becher I, Sweetman GM, Savitski MM, Bantscheff M. Higher multiplexing with isobaric mass tags enabled by ^{15}N and ^{13}C containing reporter ions and high resolution mass spectrometry. Presented at: *EuPA 2012 Scientific Congress*. Glasgow, UK, 9–12 July 2012.

Websites

101 Peptide shaker. code.google.com/p/peptide-shaker

102 Pep Tracker. www.peptracker.com

103 Host–pathogen interaction – The *C. albicans* macrophage model. www.ucm.es/info/candida

104 French Society for Electrophoresis and Proteomics Analysis. www.sfeap.fr