analytical chemistry

First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study

Claudio Iacobucci,[a,#] Christine Piotrowski,[a,#] Ruedi Aebersold,[b,c] Bruno C. Amaral,[d] Philip Andrews,[e] Katja Bernfur,[q] Christoph Borchers,[f,g,h,i] Nicolas I. Brodie,[f] James E. Bruce,[j] Yong Cao,[o] Stéphane Chaignepain,[k] Juan D. Chavez,[j] Stéphane Claverol,[l] Jürgen Cox,[m] Trisha Davis,[oo] Gianluca Degliesposti,[n] Meng-Qiu Dong,[o] Nufar Edinger,[p] Cecilia Emanuelsson,[q] Marina Gay,[r] Michael Götze,[s] Francisco Gomes-Neto,[rr] Fabio C. Gozzo,[d] Craig Gutierrez,[t] Caroline Haupt,[u] Albert J. R. Heck,[v] Franz Herzog,[w] Lan Huang,[t] Michael R. Hoopmann,[x] Nir Kalisman,[p] Oleg Klykov,[v] Zdeněk Kukačka,[y] Fan Liu,[z] Michael J. MacCoss,[aa] Karl Mechtler,[bb] Ravit Mesika,[p] Robert L. Moritz,[x] Nagarjuna Nagaraj,[cc] Victor Nesati,[dd] Ana G. C. Neves-Ferreira,[rr] Robert Ninnis,[dd] Petr Novák,[y] Francis J. O'Reilly,[ee] Matthias Pelzing,[dd] Evgeniy Petrotchenko,[f] Lolita Piersimoni,[e] Manolo Plasencia,[e] Tara Pukala,[ff] Kasper D. Rand,[gg] Juri Rappsilber,[ee,hh] Dana Reichmann,[p] Carolin Sailer,[ii] Chris P. Sarnowski,[b,jj] Richard A. Scheltema,[v] Carla Schmidt,[u] David C. Schriemer,[kk] Yi Shi,[nn] J. Mark Skehel,[n] Moriya Slavin,[p] Frank Sobott,[pp,qq] Victor Solis-Mezarino,[w] Heike Stephanowitz,[z] Florian Stengel,[ii] Christian E. Stieger,[bb] Esben Trabjerg,[gg] Michael Trnka,[ll] Marta Vilaseca,[r] Rosa Viner,[mm] Yufei Xiang,[nn] Sule Yilmaz,[m] Alex Zelter,[oo] Daniel Ziemianowicz,[kk] Alexander Leitner,*,[b] and Andrea Sinz*,[a]

[a]Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3a, 06120 Halle/Saale, Germany

[b]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, 8093 Zurich, Switzerland

[c]Faculty of Science, University of Zurich, 8006 Zurich, Switzerland

[d]Institute of Chemistry, University of Campinas, Campinas São Paulo 13083-970, Brazil

[e]Departments of Biological Chemistry, Bioinformatics, and Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

[f]University of Victoria−Genome British Columbia Proteomics Centre, Vancouver Island Technology Park, Victoria, British Columbia V8Z 7X8, Canada

[g]Department of Biochemistry and Microbiology, University of Victoria, Petch Building, Room 270d, 3800 Finnerty Road, Victoria, British Columbia V8P 5C2, Canada

[h]Gerald Bronfman Department of Oncology, Jewish General Hospital, McGill University, 3755 Côte Ste-Catherine Road, Montréal, Quebec H3T 1E2, Canada

[i]Proteomics Centre, Segal Cancer Centre, Lady Davis Institute, Jewish General Hospital, McGill University, 3755 Côte Ste-Catherine Road, Montréal, Quebec H3T 1E2, Canada

[j]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, United States

[k]CBMN, UMR 5248, CNRS, Université de Bordeaux, INP Bordeaux, Pessac 33607, France

[l]Centre de Génomique Fonctionnelle, Plateforme Protéome, Université de Bordeaux, Bordeaux33000, France

[m]Computational Systems Biochemistry Research Group, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

[n]MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, U.K.

[o]National Institute of Biological Sciences, Beijing 7 Science Park Road, ZGC Life Science Park, 102206 Beijing, China

[p]Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, Safra Campus Givat Ram, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

[q]Department of Biochemistry and Structural Biology, Center for Molecular Protein Science, Lund University, 221 00 Lund, Sweden

[r]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Baldiri Reixac 10, 08028 Barcelona, Spain

[s]Institute for Biochemistry and Biotechnology, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3a, 06120 Halle/Saale, Germany

[t]Department of Physiology & Biophysics, University of California, Irvine, California 92697, United States

[u]Interdisciplinary Research Center HALOmem, Institute for Biochemistry and Biotechnology, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3a, 06120 Halle/Saale, Germany

[v]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht and Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

[w]Gene Center Munich, Department of Biochemistry, Faculty of Chemistry and Pharmacy, Ludwig Maximilians University of Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

[x]Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98109, United States

[y]Institute of Microbiology, BIOCEV, Prumyslova 595, 252 50 Vestec, Czech Republic

[z]Leibniz Institute of Molecular Pharmacology (FMP), Robert-Rössle-Strasse 10, 13125 Berlin, Germany

[aa]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, United States

[bb]Protein Chemistry Facility, Research Institute of Molecular Pathology (IMP) and Institute of Molecular Biotechnology (IMBA), Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[cc]Biochemistry Core Facility, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

[dd]Analytical Biochemistry, CSL Limited, Bio21 Institute, 30 Flemington Road, 3010 Parkville, Melbourne, Australia

[ee]Chair of Bioanalytics, Institute of Biotechnology Technische Universität Berlin, 13355 Berlin, Germany

[ff]Discipline of Chemistry, Faculty of Sciences, University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia

[gg]Department of Pharmacy, University of Copenhagen, 2100 Copenhagen, Denmark

[hh]Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, EH9 3BF Edinburgh, U.K.

[ii]University of Konstanz, Department of Biology, Universitätsstrasse 10, 78457 Konstanz, Germany

[jj]PhD Program in Systems Biology, University of Zurich and ETH Zurich, 8092 Zurich, Switzerland

[kk]Department of Biochemistry & Molecular Biology, Robson DNA Science Centre, University of Calgary, 3330 Hospital Drive North West, Calgary, Alberta T2N 4N1, Canada

[ll]UCSF Mass Spectrometry Facility, Genentech Hall, 600 16th Street, San Francisco, California 94158, United States

[mm]Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose, California 95134, United States

[nn]Department of Cell Biology, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania 15213, United States

[oo]Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States

[pp]Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

[qq]The Astbury Centre for Structural Molecular Biology and School of Molecular and Cellular Biology, University of Leeds, LS2 9JT Leeds, U.K.

[rr]Laboratory of Toxinology, Oswaldo Cruz Institute, Fiocruz, Avenida Brasil 4365 (Moorish Castle), Manguinhos, Rio de Janeiro, Rio de Janeiro 21040-900, Brazil

ⓈSupporting Information

ABSTRACT: The number of publications in the field of chemical cross-linking combined with mass spectrometry (XL-MS) to derive constraints for protein three-dimensional structure modeling and to probe protein−protein interactions has increased during the last years. As the technique is now becoming routine for in vitro and in vivo applications in proteomics and structural biology there is a pressing need to define protocols as well as data analysis and reporting formats. Such consensus formats should become accepted in the field and be shown to lead to reproducible results. This first, *continued...*
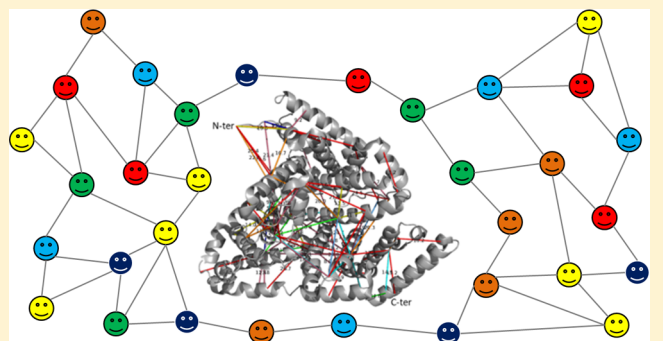
community-based harmonization study on XL-MS is based on the results of 32 groups participating worldwide. The aim of this paper is to summarize the status quo of XL-MS and to compare and evaluate existing cross-linking strategies. Our study therefore builds the framework for establishing best practice guidelines to conduct cross-linking experiments, perform data analysis, and define reporting formats with the ultimate goal of assisting scientists to generate accurate and reproducible XL-MS results.

Mass spectrometry (MS) is becoming increasingly popular in the field of structural biology, with great implications for solving important biological questions. A central technique in structural MS is chemical cross-linking combined with MS (XL-MS). Since 2000, XL-MS and computational modeling has advanced from investigating three-dimensional structures of isolated proteins to deciphering protein interaction networks.[1−4] In the field of integrated structure analysis, XL-MS is often used in conjunction with cryo-electron microscopy. As the chemical XL-MS approach allows the capture of transient and weak interactions, it is now becoming a routine technique for unraveling protein interaction networks in their natural cellular environment.[5] The knowledge obtained will significantly advance our understanding of the structure of functional complexes, the topology of cellular networks and molecular details underlying human pathologies.

Briefly, the XL-MS approach relies on adding a chemical reagent to a protein solution connecting two functional groups of amino acid side chains. Cross-linker molecules consist of two reactive groups that are separated via a spacer of defined length that allow to derive distance information on a protein or a protein assembly. The cross-linked residues are usually identified after enzymatic digestion of the covalently connected protein(s) using LC/ESI-MS/MS (liquid chromatography/electrospray ionization-tandem mass spectrometry) and the resulting fragment ion spectra are computationally assigned to the cross-linked peptides. The distance constraints imposed by the chemical cross-linker on the protein's tertiary structure serve as a basis for subsequent computational modeling studies to derive three-dimensional structural models (Scheme S1). XL-MS can be applied to both proteins and protein complexes and in the case of protein assemblies, the distance constraints can be used to map the subunit topology. XL-MS is now increasingly being used for deriving protein−protein inter-action maps, both in vitro and in vivo, where interacting proteins are covalently connected by the cross-linking reaction.[6−13]

The wide acceptance of XL-MS by the proteomics and structural biology communities reflects the increasing importance of cross-linking data for elucidating protein structures and protein−protein interactions. However, the growth of the user base brings about challenges of its own: Even a relatively superficial glance at the literature shows a huge diversity of cross-linkers, experimental workflows, and computational pipelines. Moreover, the information provided in scientific research articles that contain cross-linking data can range from being quite detailed to very brief.

The heterogeneity of cross-linking protocols has mainly emerged from the use of different cross-linking chemistries and different designs of the corresponding cross-linker (e.g., noncleavable/cleavable, isotope-coded, or affinity-tagged re-agents). This, in turn, necessitated individual software solutions specifically tailored to the analysis of data from the experimental workflow. The most common database search engines used in proteomics are not directly suitable for interpreting mass spectra from cross-linked peptides. There-fore, the majority of computational solutions have emerged from laboratories that pioneered the application of XL-MS and created tools specifically tailored for the analysis of cross-linked peptides. Together with a current lack of formal or even informal reporting standards, the present state of XL-MS may confuse researchers that are interested in interpreting results from XL-MS studies or in adopting the technology. Currently, it is not clear which strategies are most suitable in general or for a particular application, which makes it challenging to objectively compare results obtained by different groups.

Certainly, the challenges summarized above resemble those of other disciplines. In particular, scientists active in "conven-tional" proteomics research have tried to address the very same issues over the past decade. Interlaboratory and software comparison studies have been performed for different experimental strategies, including data-dependent acquisi-tion,[14] selected reaction monitoring,[15−18] and most recently, data-independent acquisition.[19,20] In addition, regular com-parative studies have been organized by the Association of Biomolecular Resource Facilities (ABRF; https://abrf.org/research-group/proteomics-research-group-prg and https://abrf.org/research-group/proteomics-standards-research-group-sprg). Together, these studies revealed limitations in commonly used experimental and computational workflows, but on the other hand also provided evidence for the robustness of a particular technique when implemented in different laboratories according to standard operating procedures.

Standardized file formats and reporting guidelines for proteomics have been developed under the auspices of the Proteomics Standards Initiative (PSI) of the Human Proteome Organization (http://www.psidev.info).[21] For example, as far back as 2007, the first recommendations for minimum reporting standards in proteomics (Minimum Information About a Proteomics Experiment, MIAPE) have been made,[22] which have been followed by detailed guidelines of several proteomics journals. PSI has also formalized open-file formats, such as the mzML format for raw MS data[23] and the mzIdentML format for protein identifications.[24] Such guide-lines and open data formats have also led to an increase in the deposition of proteomics data in open data repositories such as the PRoteomics IDEntifications (PRIDE) archive, hosted by the European Bioinformatics Institute (https://www.ebi.ac.uk/pride/archive/),[25] via the ProteomeXchange initiative (https://www.proteomexchange.org).[26]

Initiatives for establishing standards and recommendations of best practices within structural MS techniques, ion mobility-MS (https://chemrxiv.org/articles/Recommendations_for_Reporting_Ion_Mobility_Mass_Spectrometry_Measurements/7072070), hydrogen/deuterium exchange (manuscripts in preparation), and native MS are or have recently emerged. Likewise, there is also a clear need for the objective assessment of the methods and reporting standards within the field of XL-MS. For this purpose, several researchers active in the field of XL-MS decided to start a community-

organized effort with the goal of providing a first overview of common procedures in XL-MS to generate the basis for best practices in the field.

In this first interlaboratory effort, 32 groups worldwide contributed, delivering a total of 58 cross-linking data sets. The data reflect the great diversity of experimental and computational strategies employed, and to our knowledge, this is the first comprehensive study with the aim to harmonize the XL-MS field.

■ RESULTS

Study Design. We opted for a simple study design to encourage participation from as many laboratories as possible, including those with currently only little experience in XL-MS. Invitations were sent out to research groups known to be active in the field from their published work and to attendants of the Symposium of Structural Proteomics (SSP, http://www.structuralproteomics.net/) meeting series. The guidelines were kept quite simple, and each participant was provided with a template spreadsheet to document their method and report their results (Supporting Information). Bovine serum albumin (BSA), a protein with a molecular weight of ~66 kDa, was selected as the study system. We requested that a certain product from a widely available supplier should be used, and it was specified to use a BSA concentration of 10 μM. Apart from these restrictions, we left the contributing laboratories full freedom to choose the experimental and computational strategies of their choice. This included, among other parameters, flexibility regarding the choice of cross-linking reagent and its concentration, buffer composition and pH, reaction time and temperature, post-cross-linking sample processing (digestion protocol, optional fractionation, and enrichment of cross-linked products), conditions for LC/MS analysis, and data analysis procedures (choice of software, search parameters, validation of the results). In short, we expected that participants would use the typical XL-MS workflows established in their laboratories. The protocols used by the individual participating laboratories were collected and analyzed in the Sinz lab and are summarized in the Supporting Information.

For data analysis, we provided the amino acid sequence of mature BSA after cleavage of the signal peptide and propeptide sequences (residues 25−607 of the UniProt entry P02769, https://www.uniprot.org/uniprot/P02769) to ensure a uniform numbering scheme. Finally, we encouraged participants to perform at least three replicates. As mentioned above, we provided a template spreadsheet (Supporting Information) that needed to be completed by the participants before a data set would be considered for inclusion in the detailed assessment of the results. An overview of the data sets provided by different laboratories is presented in Figure 1.

Protein System. BSA was selected as model protein for this study as it is a globular and stable protein that is readily available at low cost. Moreover, the three-dimensional structure of BSA is well-known, and we selected the Protein Data Bank entry 4F5S (https://www.rcsb.org/structure/4F5S) for further interpretation of the results. As BSA possesses a tendency toward forming dimers, this has to be considered when interpreting the results (see also below).

Cross-Linking Reagents. As outlined above, the participants of this study were free to choose the cross-linking principle(s) on their own (Table S1, Supporting Information). The majority of groups decided to use noncleavable,
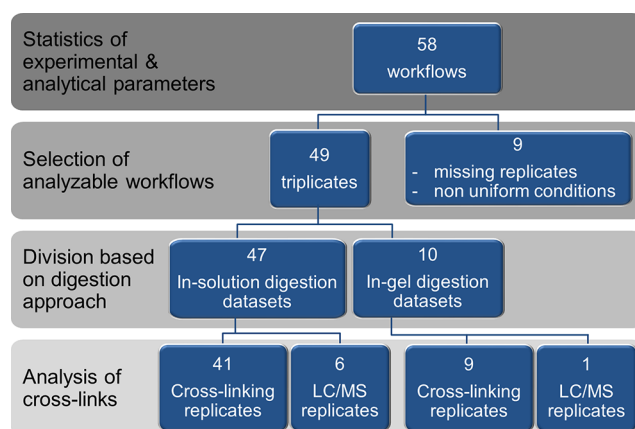

Figure 1. Overview of data sets provided by the participants of this study: 32 groups participated in this study, yielding 58 separate cross-linking workflows. Nine data sets had to be excluded due to missing replicates and nonuniform conditions, resulting in a total of 49 data sets that were further considered. Several workflows contain both in-solution (47 samples) as well as in-gel digestion (10 samples) as processing methods. The samples were considered only once during a workflow analysis.
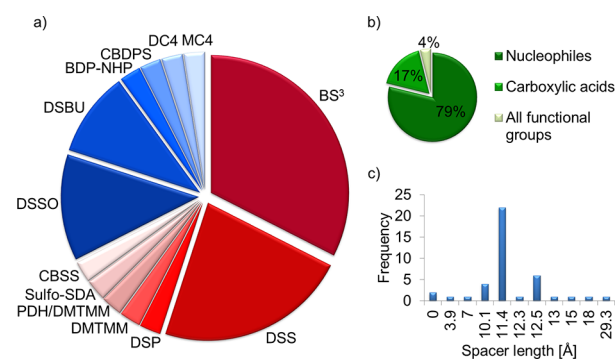

Figure 2. (a) Cross-linking reagents used in this study; noncleavable cross-linkers are presented in red, MS-cleavable cross-linkers are shown in blue, (b) reactivity, and (c) spacer length. The cross-linkers used in this study are BS[3] (bis(sulfosuccinimidyl)suberate, DSS (disuccinimidylsuberate), DSP (dithiobis(succinimidylpropionate)), DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride) with and without PDH (pimelic acid dihydrazide), sulfo-SDA (sulfosuccinimidyl 4,4′-azipentanoate), CBSS (carboxybenzophenone sulfosuccinimide), DSSO (disuccinimidylsulfoxide), DSBU (disuccinimidyldibutyric urea), BDP-NHP (N-hydroxyphthalamide ester of biotin aspartate proline), CBDPS (cyanurbiotindimercaptopropionyl succinimide), DC4 (1,4-bis(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium), and MC4 (N,N′-bis(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-morpholine).

homobifunctional, amine-reactive N-hydroxysuccinimide (NHS) cross-linkers, i.e., bis(sulfosuccinimidyl)suberate (BS[3]) or disuccinimidylsuberate (DSS) (Figure 2a). Both cross-linkers only differ by a sulfonic acid group that is incorporated for increased water solubility and bridge a distance of 11.4 Å, resulting in Cα−Cα distances of ~27 Å to be cross-linked.[27] MS-cleavable cross-linkers, such as disuccinimidylsulfoxide (DSSO) and disuccinimidyldibutyric urea (DSBU), are increasingly being used as they allow a targeted identification of cross-linked product based on characteristic reporter ions generated during MS/MS experiments. MS-cleavability as a cross-linker feature is essential to

reduce the search space in conducting proteome-wide cross-linking studies. The vast majority of cross-linkers used herein target amine groups in proteins, i.e., lysine side chains, while carboxylic acid groups, such as aspartic and glutamic acid residues, are less frequently targeted (Figure 2b). The main spacer lengths of the cross-linkers are determined by the three
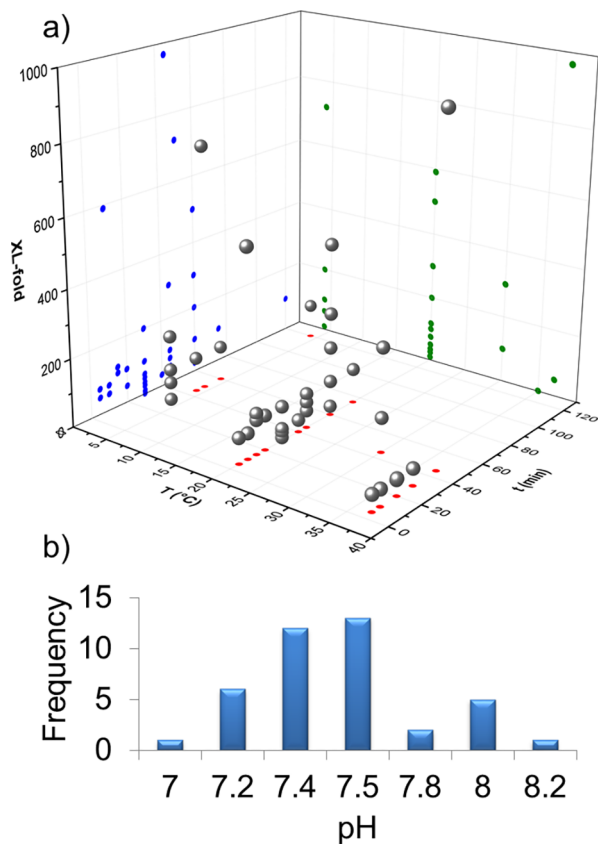

Figure 3. (a) Time, temperature, and cross-linker excess (XL-fold) were set as variable parameters, presented as gray spheres. The colored dots are projections of the 3D space onto 2D planes. (b) pH values of the cross-linking reactions ranged between 7.0 and 8.2.

most abundant cross-linkers used in this study, BS[3] and DSS (both 11.4 Å), DSBU (12.5 Å), and DSSO (10.1 Å) (Figure 2c).

Reaction Conditions. The reaction conditions were also kept completely open to the participants, including cross-linking reaction time, temperature, cross-linker excess, and pH value of the cross-linking solution (Figure 3). Not surprisingly, the pH value of the cross-linking reaction mixture was kept around pH 7.4 to 7.5 in the majority of experiments in order to resemble the physiological pH situation. A pH value of 8.0 that was also used in some experiments has the advantage of enhancing the reactivity of NHS esters with nucleophiles. The temperature was kept to 20, 25, or 37 °C in the majority of experiments, with lower temperature being applied only by a few groups. For BSA, a temperature of 37 °C certainly does not present a problem as it is a stable, globular protein, but for delicate and unstable proteins one should take care to conduct the cross-linking reaction at lower temperatures.

Instrument Platforms and Settings Used to Generate XL-MS Data. The overwhelming majority of cross-linking data were generated on orbitrap mass spectrometers (Figure 4). Only two FTICR (SolariX and Velos FTICR) mass
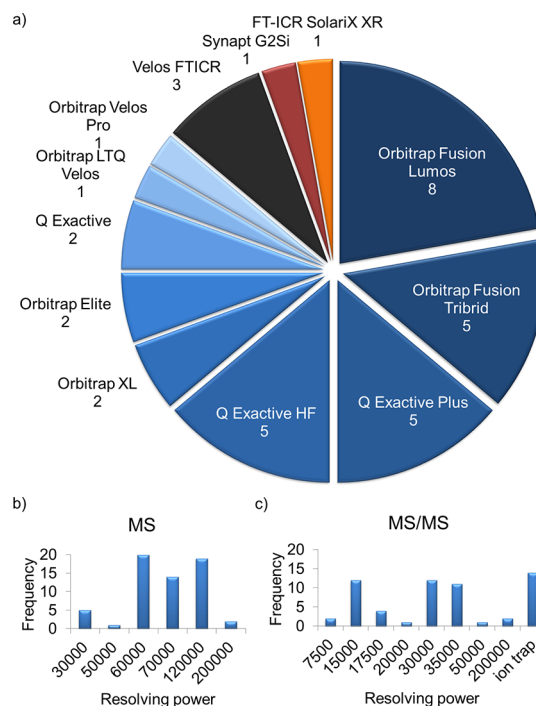

Figure 4. LC/MS/MS conditions applied. (a) MS instrumentation, (b) MS resolving power, and (c) MS/MS resolving power. Resolving power is defined at m/z 200 for orbitrap instruments, while for ICR instruments it is defined at m/z 400. Please note that several research groups generated data sets with different instruments and settings.

spectrometers and one Q-TOF (Synapt G2 SI) instrument were employed (Figure 4a). All groups used LC/ESI-MS/MS analysis, applying for most experiments a resolving power of 60 000 or 120 000 (at m/z 200 or 400, as specified by the manufacturer Thermo Fisher Scientific for orbitrap instru-
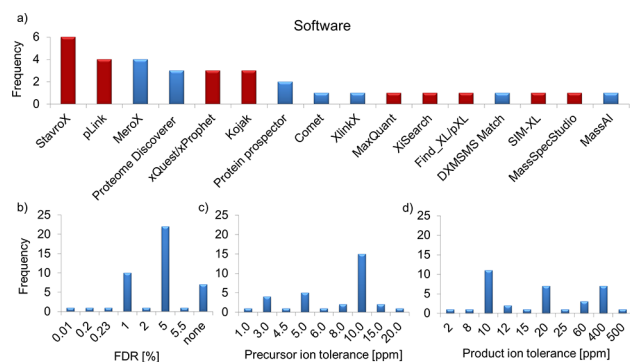

Figure 5. (a) Software tools used in this study (a complete summary is found in Table S2, Supporting Information). Red bars indicate that the software is applicable only for noncleavable cross-linkers; blue bars indicate that the software can be used for MS-cleavable cross-linkers. (b) False discovery rates. (c) Mass tolerance MS. (d) Mass tolerance MS/MS. For the Proteome Discoverer, data analysis was performed using the XlinkX software node.

ments) (Figure 4b). For MS/MS experiments, a resolving power of 15 000 or 30 000 was employed in most cases (Figure 4c). Details on enrichment of cross-linked species, considered charge states, fragmentation methods, and MS[3] resolution are presented in the Supporting Information (Figure S1).

Data Analysis and Validation Strategies. Strategies for data analysis were highly diverse (Figure 5), reflecting the variety in the XL-MS field where nearly every group possesses their own software tools tailored to fit their specific needs. This enormous variety is currently one of the most critical issues in XL-MS, and we consider it as an important contribution of this study to reflect this diversity. The false discovery rate (FDR) plays an important role in this context, and from this study it arose that most of the groups apply an FDR of 5% (Figure 5b). Manual validation of the cross-links was performed for 66% of the experiments, while in 34%, the data sets were not manually checked. It is important to note that a mechanism to control the FDR should exist in the software; although proper FDR control is not trivial for small search spaces, manual validation strategies might be especially beneficial in such cases. Some strategies provide additional layers of evidence that can be used to better control the error rate. For example, isotope-coded, noncleavable linkers provide two independent measures of precursor and fragment masses and charge state information for fragments independent of MS resolution; MS-cleavable linkers provide three layers of information: intact precursors, released fragments corresponding to intact peptide chains, and fragments thereof. In the absence of such strategies, we recommend that preferentially both, MS and MS/MS data, should be recorded with high mass accuracy to rule out a false assignment of cross-linked products. Clearly, some of these effects will only become apparent for samples of higher complexity.

Identified Cross-Links. As we left it to the individual participants whether to use in-solution or in-gel digestion as the workup method before LC/MS/MS analysis, 47 data sets were generated by in-solution digestion, while 10 samples originated from in-gel digestion (Figure 1). As already mentioned, BSA has a tendency to form dimers, which somewhat complicates data analysis. In case only the BSA monomer band is used for in-gel digestion and subsequent generation of the cross-linking data set, one can definitely rule out that cross-links are in fact representing intermolecular interactions between two BSA molecules. On the other hand, during the in-gel digestion procedure cross-links might get lost, resulting in an overall lower number of cross-linked products.

Another aspect regards the reaction sites that were considered during data analysis. Usually, NHS esters, such as the mainly used cross-linkers BS[3], DSS, DSBU, and DSSO, will react with lysine, but they also exhibit a significant reactivity toward serine, threonine, and tyrosine. The pH used for conducting the cross-linking reaction plays a significant role as amine reactivity is increased at higher pH values. Some participants considered only Lys–Lys cross-links and neglected the side-reactivity of NHS esters with hydroxy group-containing amino acids. In this study, it became apparent that Ser, Thr, and Tyr account for ~30% of cross-linking sites (Supporting Information, Figure S2). The reactivity of Ser, Thr, and Tyr residues obviously depends on the reaction conditions (cross-linker, pH value of the solution) as well as local pK_a value. It is not practicable to consider Lys, Ser, Thr, and Tyr when analyzing very complex systems, such as complete proteomes. Therefore, we suggest as a compromise to consider for whole proteome samples only lysine as the reactive sites of NHS ester cross-linkers, while for single proteins or proteins assemblies, Lys, Ser, Thr, and Tyr might be taken into account.
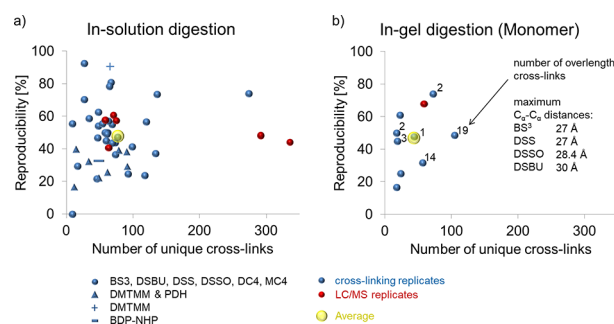

Figure 6. Number of BSA cross-links identified. The numbers of cross-links are plotted for (a) in-solution and (b) in-gel digestion workflows. The different cross-linkers are shown as symbols; abbreviations of the cross-linkers are according to Figure 2. The maximum distances are given for each cross-linker, indicating the number of overlength cross-links. Every point is a sum of three replicate measurements; replicates of the entire experiment are shown in blue, and replicates of the LC/MS analyses are shown in red; the average number and reproducibility of unique cross-links are shown in yellow.

Figure 6 provides an overview about the reproducibility of results obtained with the individual workflows of the participants. For in-solution digestion workflows, the average number of unique cross-links in BSA is 78, while for in-gel digestion workflows using only the monomeric BSA band, the average number is 44. The term "cross-link" refers to the specific amino acid residues that are connected, irrespective of different peptide sequences due to missed cleavage sites or modifications. The majority of participating laboratories came up with similar numbers of unique cross-links, independently of the cross-linking conditions used (Figure 6a). Three cross-linking workflows however recorded a significantly higher number of cross-links (between 260 and 350). The reason could be a false consideration of cross-links from BSA dimers that in some preparations might have been a dominating species due to inappropriate sample treatment. For in-gel digestion workflows, up to 19 overlength cross-links were reported in one data set, which could represent false-positives due to partial unfolding as only the monomeric form of BSA was considered in these samples (Figure 6b).

A more detailed inspection of the unique cross-links revealed highly interesting insights: Data sets created from amine-reactive cross-linkers (BS[3], DSBU, DSS, DSSO, DC4, MC4, CBDPS) using an in-solution digestion workflow yielded a total of 1066 unique cross-links. A complete list of unique cross-links, identified with cross-linkers reacting with nucleo-philes (amine and hydroxy groups) and sorted by their reproducibility, is provided as separate file in the Supporting Information. In total, 601 of 1066 unique cross-links (56%) were however identified in only one single data set (Figure 7). This indicates an overall low reproducibility of cross-linking results. The curve in Figure 7a shows that the number of unique cross-links identified is inversely proportional to the reproducibility of cross-links in the data sets (coefficient of proportionality $\simeq -1$). If the reproducibility across the data sets is higher than 20%, the effect of including more data sets, different reaction conditions, and analytical parameters determines a linear increment of the number of cross-link identifications. The intercept with the y-axis of the resulting interpolated linear curves indicates the putative number of cross-links in BSA to be between 73 and 88 (Figure 7b). This
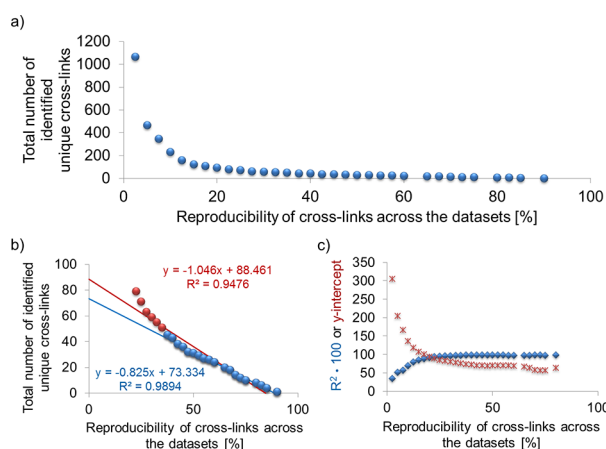
Figure 7. Comparison of unique cross-links. "Cross-link" denotes the unique amino acid residues that are connected by homobifunctional, amine-reactive cross-linkers. (a) Number of cross-links with respect to their reproducibility among the data sets. (b) Linear extrapolation of all (red) or a linear subset (blue) of cross-links resulted in a maximum cross-linking number between 73 and 88. (c) Plot summarizes the intercepts with the y-axis (red) and the correlation coefficient $\times$ 100 (blue) of the respective linear extrapolations of part a. The linear extrapolation was calculated as shown in part b by successively removing the data points starting from the lowest reproducible value.

value is very close to the average number of cross-links found (78 cross-links per data set for in-solution digestion workflows, Figure 6a). In Figure 7c, the dependence of the linear correlation on the reproducibility of cross-links identified is indicated. This indicates that a linear correlation only exists for highly reproducible cross-links.

Cross-Links Identified from In-Gel Digested BSA Monomer Band. We mapped cross-links in the monomer band of BSA using *in-gel* digestion (in total 10 data sets) into the published 3D structure of BSA (PDB entry 4F5S). For this, a statistical analysis could be performed for homobifunctional, amine-reactive linkers considered only for this type of cross-linker. Only cross-links identified in at least two independent experiments are presented (Figure S3). A total of 30 out of 230 cross-links exceeds the maximum length of 30 Å for the cross-linkers employed in this study. These overlength cross-links either originate from a false assignment or by applying nonsuitable experimental conditions. Strikingly, 29 of these overlength cross-links were identified in one single experiment only. Cross-links that were identified in at least two independent experiments show one overlength link, while cross-links found in at least three independent experiments all fall within the given distance limit of 30 Å (Figure S4). As guideline for testing cross-linking workflows, we provide a list of cross-links that were identified in at least two independent experiments from in-gel digestion of the BSA monomer band (Table S3, Supporting Information).

Monomer−Dimer Equilibrium of BSA. BSA exists in a monomer−dimer equilibrium, which may give rise to ambiguities in the identification of intra- and intermolecular cross-links. To address this issue, we performed additional experiments with four concentrations of BSA (10, 5, 1, and 0.5 μM). Strikingly, the number of overlength cross-links was very low (only 1 or 2). Moreover, the numbers of overlength cross-links were similar for all four BSA concentrations used (Table S4, Supporting Information). This clearly indicates that a BSA

concentration of 10 μM, as chosen for this study, is suitable for conducting cross-linking MS experiments.

Comparison of Data Acquisition and Analysis Strategies from One Participating Laboratory. Because most of the data in this study have been generated in different laboratories, differences in instrumentation and in the software used for data analysis make a direct comparison of selected results difficult. However, we used a subset of the data generated in a single laboratory to study the effect of the type of mass spectrometer and of different search settings on the outcome for a relatively simple model system, such as BSA (see Supporting Information).

■ DISCUSSION

This first community-based cross-linking study reflects the high diversity of XL-MS workflows that are currently employed in different laboratories worldwide. However, it also became apparent that independent of the workflow used, the results obtained are to some degree comparable. For beginners in the field, we suggest to use BSA as an initial study system and compare the outcome to the results obtained herein. As a guideline, the number of cross-links expected for BSA should be ∼80 for an in-solution workflow, considering cross-links of the monomer and the dimer. Not unexpectedly, our study did not reveal the optimum experimental protocol or software to be used in any and all projects. The applications of XL-MS are just too diverse so that no single cross-linker, instrument, or software tool is expected to be preferable for all scenarios, ranging from single protein (as used in this work) to whole-cell cross-linking. There are also clear interdependencies between the type of cross-linker (cleavable, noncleavable) and the software that can be applied to process such data as well as between instrument type and software as not all fragmentation methods or other MS platform-dependent features may be supported.

As discussed above, XL-MS has become an essential part of many structural proteomics studies but is also a key element in integrative structural biology projects. In such interdisciplinary work, XL data may only be a small "puzzle piece" that is combined with other experimental data provided by methods such as electron microscopy, X-ray crystallography, NMR spectroscopy, small-angle X-ray scattering, together with computational modeling. Details about how experiments were carried out, how the data were processed, and how error rates were assessed are often missing from the publication, making it difficult for reviewers and readers to assess the reliability and credibility of the results. We therefore recommend that appropriate consideration should be given to the method section of all XL-MS publications by providing all necessary experimental and computational details. Our reporting template could serve as a starting point for the "minimum information about a cross-linking experiment" that should be included in research articles containing XL-MS data. This template is included in the Supporting Information for all XL-MS data reports. Sufficient information needs to be provided, irrespective of the relative contribution of the cross-linking experiments to a specific project. This will also facilitate the cross-referencing of XL-MS data in integrative structural biology projects, for example, in the dedicated PDB prototype archive, PDB-Dev.[28]

Data deposition to a proteomics repository, such as PRIDE, is encouraged, as the paucity of available data sets do not assist the field in validation, methods evaluation, and workflow

Table 1. Cross-Linking Mass Spectrometry Guidelines (Guidelines 1 and 2 Are Derived from the Results Shown in Figure 5)

no.	topic	description
1	FDR	A mechanism to control the FDR should exist in the software used for cross-link identification. The FDR algorithm has to be described in detail. For small search spaces, manual validation strategies might be beneficial.
2	mass accuracy	MS and MS/MS data should be recorded and analyzed with high mass accuracy to reduce false assignments of cross-linked products, or multiple lines of evidence from isotope labeling or cleavable linkers should be obtained.
3	experimental details	Provide all experimental and computational details. The reporting template (Supporting Information) comprises the "minimum information of a cross-linking experiment" that should be included in research articles containing XL-MS data.
4	data deposition	Deposit raw MS files together with a description of their content and the reporting template to a proteomics repository, such as PRIDE.
5	visualization of cross-linked proteins	Perform SDS-PAGE analysis to evaluate the cross-linking performance under the employed experimental conditions. Check for possible high-molecular weight aggregates.
6	cross-linker selectivity	Consider only lysine and the N-terminus as reactive sites of amine-reactive cross-linkers for whole proteome samples. For single proteins or large protein assemblies, consider lysine, N-terminus, serine, threonine, and tyrosine as reactive sites.
7	BSA cross-links	Approximately 80 cross-links can be expected for cross-linking of BSA using homobifunctional amine-reactive cross-linkers and an in-solution digestion workflow.

quality. It should be noted that not all data sets assigned to the cross-linking category in PRIDE originate from genuine XL-MS experiments (in the sense that cross-linking sites were identified) but also contain data from experiments that used cross-linking for the stabilization of complexes. The low uptake of data deposition may in part be due to the specific nature of XL-MS data. For a "complete" submission to ProteomeXchange, allowing a complete integration of search results and assignment of a Digital Object Identifier, the reported results need to be compliant with a PSI format, such as mzIdentML. Although the most recent version of mzIdentML (version 1.2) includes support for some XL-MS strategies, such a proteomics-centered format cannot easily consider all possible workflows, and few dedicated cross-linking search engines offer mzIdentML-compliant export at this point. Nevertheless, even a "partial" submission will make the raw MS data and results available in a user-specified format for download and reuse by interested researchers.

Additional studies that cover a wider range of sample types, such as large multiprotein assemblies or even whole proteomes, will be required to obtain a better understanding of the benefits and drawbacks of different experimental workflows. However, we believe that this first community-based study serves as the starting point for further initiatives in this direction and encourages the adoption of consistent reporting and data sharing guidelines in XL-MS. We would like to invite interested parties to participate in the discussion to expand the growing XL-MS community.

CONCLUSION AND GUIDELINES

Although XL-MS is becoming routine for in vitro and in vivo applications in proteomics and structural biology, this harmonization initiative unveiled a great variety in the cross-links identified by participating groups, even for the single protein BSA. This underlines the need for establishing generally accepted XL-MS protocols as well as data analysis and reporting formats. This interlaboratory study on XL-MS represents the first effort of the community toward establishing endorsed and transparent good practice guidelines for performing and reporting XL-MS experiments. This study also serves as test for all laboratories to evaluate the quality of their XL-MS workflows and will aid in improving eventual weaknesses. In summary, seven guidelines were deduced from this study as framework for conducting XL-MS experiments as detailed in Table 1.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.9b00658.

General workflow of cross-linking mass spectrometry; details on enrichment of cross-linked species, considered charge states, fragmentation methods, and MS^3 resolution; influence of cross-linking sites considered in data analysis; cross-links with homobifunctional amine-reactive reagents found and identified in the monomer band of BSA using *in-gel* digestion; list of cross-linking reagents and software used in this study; list of unique cross-links identified with homobifunctional, amine-reactive reagents after *in-gel* digestion of the BSA monomer band; unique cross-links identified at different BSA concentrations; and comparison of data acquisition and analysis strategies from one participating laboratory (PDF)

Example of reporting template (XLSX)

Complete list of unique cross-links identified with cross-linkers reacting with nucleophiles (amine and hydroxy groups) and sorted by their reproducibility (XLSX)

Unique cross-links and number of identifications (PDF)

Unique cross-links identified at different BSA concentrations (XLSX)

Reporting templates for all XL-MS data reports (ZIP)

AUTHOR INFORMATION

Corresponding Authors
*E-mail: andrea.sinz@pharmazie.uni-halle.de.
*E-mail: leitner@imsb.biol.ethz.ch.

ORCID Ⓞ
Christoph Borchers: 0000-0003-2394-6512
James E. Bruce: 0000-0001-6441-6089
Jürgen Cox: 0000-0001-8597-205X
Cecilia Emanuelsson: 0000-0001-8762-477X
Fabio C. Gozzo: 0000-0002-5270-4427
Albert J. R. Heck: 0000-0002-2405-4404
Lan Huang: 0000-0002-3140-4687
Michael R. Hoopmann: 0000-0001-7029-7792
Oleg Klykov: 0000-0003-4401-9400
Karl Mechtler: 0000-0002-3392-9946
Robert L. Moritz: 0000-0002-3216-9447
Petr Novák: 0000-0001-8688-529X

Tara Pukala: 0000-0001-7391-1436
Kasper D. Rand: 0000-0002-6337-5489
Juri Rappsilber: 0000-0001-5999-1310
Richard A. Scheltema: 0000-0002-1668-0253
Carla Schmidt: 0000-0001-9410-1424
David C. Schriemer: 0000-0002-5202-1618
Alex Zelter: 0000-0002-5331-0577
Alexander Leitner: 0000-0003-4126-0725
Andrea Sinz: 0000-0003-1521-4899

Author Contributions

[#]C.I. and C.P. contributed equally. The study was designed by A.L., C.I., and A.S., and all authors contributed the data. C.I., C.P., A.L., and A.S. performed data analysis. S.F. coordinated the COST Action BM1403.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (2000), 5802−5806.

(2) Rappsilber, J.; Siniossoglou, S.; Hurt, E. C.; Mann, M. *Anal. Chem.* **2000**, *72*, 267−275.

(3) Bennett, K. L.; Kussmann, M.; Björk, P.; Godzwon, M.; Mikkelsen, M.; Sørensen, P.; Roepstorff, P. *Protein Sci.* **2000**, *9*, 1503−1518.

(4) Sinz, A.; Wang, K. *Biochemistry* **2001**, *40*, 7903−7913.

(5) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311−318.

(6) Leitner, A.; Walzthoeni, T.; Aebersold, R. *Nat. Protoc.* **2014**, *9*, 120.

(7) Schmidt, C.; Robinson, C. V. *Nat. Protoc.* **2014**, *9*, 2224.

(8) Lima, D. B.; Melchior, J. T.; Morris, J.; Barbosa, V. C.; Chamot-Rooke, J.; Fioramonte, M.; Souza, T. A. C. B.; Fischer, J. S. G.; Gozzo, F. C.; Carvalho, P. C.; Davidson, W. S. *Nat. Protoc.* **2018**, *13*, 431−458.

(9) Orbán-Németh, Z.; Beveridge, R.; Hollenstein, D. M.; Rampler, E.; Stranzl, T.; Hudecz, O.; Doblmann, J.; Schlögelhofer, P.; Mechtler, K. *Nat. Protoc.* **2018**, *13*, 478−494.

(10) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. *Nat. Commun.* **2017**, *8*, 15473.

(11) Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J.; Scheltema, R. A. *Nat. Protoc.* **2018**, *13*, 2964−2990.

(12) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. *Nat. Protoc.* **2018**, *13*, 2864−2889.

(13) Chen, Z. A.; Rappsilber, J. *Nat. Protoc.* **2019**, *14*, 171−201.

(14) Bell, A. W.; Deutsch, E. W.; Au, C. E.; Kearney, R. E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J. J. *Nat. Methods* **2009**, *6*, 423.

(15) Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; et al. *Nat. Biotechnol.* **2009**, *27*, 633.

(16) Abbatiello, S. E.; Schilling, B.; Mani, D. R.; Zimmerman, L. J.; Hall, S. C.; MacLean, B.; Albertolle, M.; Allen, S.; Burgess, M.; Cusack, M. P.; et al. *Mol. Cell. Proteomics* **2015**, *14*, 2357−2374.

(17) Kennedy, J. J.; Abbatiello, S. E.; Kim, K.; Yan, P.; Whiteaker, J. R.; Lin, C.; Kim, J. S.; Zhang, Y.; Wang, X.; Ivey, R. G.; et al. *Nat. Methods* **2014**, *11*, 149.

(18) Vialas, V.; Colomé-Calls, N.; Abian, J.; Aloria, K.; Alvarez-Llamas, G.; Antúnez, O.; Arizmendi, J. M.; Azkargorta, M.; Barceló-Batllori, S.; Barderas, M. G.; et al. *J. Proteomics* **2017**, *152*, 138−149.

(19) Navarro, P.; Kuharev, J.; Gillet, L. C.; Bernhardt, O. M.; MacLean, B.; Röst, H. L.; Tate, S. A.; Tsou, C. C.; Reiter, L.; Distler, U.; et al. *Nat. Biotechnol.* **2016**, *34* (2016), 1130.

(20) Collins, B. C.; Hunter, C. L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S. L.; Chan, D. W.; Gibson, B. W.; Gingras, A. C.; Held, J. M.; et al. *Nat. Commun.* **2017**, *8*, 291.

(21) Deutsch, E. W.; Orchard, S.; Binz, P. A.; Bittremieux, W.; Eisenacher, M.; Hermjakob, H.; Kawano, S.; Lam, H.; Mayer, G.; Menschaert, G.; et al. *J. Proteome Res.* **2017**, *16*, 4288−4298.

(22) Taylor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P. A.; Julian, R. K., Jr.; Jones, A. R.; Zhu, W.; Apweiler, R.; Aebersold, R.; Deutsch, E. W.; et al. *Nat. Biotechnol.* **2007**, *25*, 887.

(23) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; et al. *Mol. Cell. Proteomics* **2011**, *10*, R110.000133.

(24) Vizcaíno, J. A.; Mayer, G.; Perkins, S.; Barsnes, H.; Vaudel, M.; Perez-Riverol, Y.; Ternent, T.; Uszkoreit, J.; Eisenacher, M.; Fischer, L.; et al. *Mol. Cell. Proteomics* **2017**, *16*, 1275−1285.

(25) Vizcaíno, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. *Nucleic Acids Res.* **2016**, *44*, D447−D456.

(26) Deutsch, E. W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, A.; Ternent, T.; Campbell, D. S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; et al. *Nucleic Acids Res.* **2017**, *45*, D1100−D1106.

(27) Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. *Protein Sci.* **2014**, *23*, 747−759.

(28) Vallat, B.; Webb, B.; Westbrook, J. D.; Sali, A.; Berman, H. M. *Structure* **2018**, *26*, 894−904.