An integrated workflow for cross-linking/mass spectrometry

Marta L. Mendes[1#], Lutz Fischer[2#], Zhuo A. Chen[1], Marta Barbon[3,4], Francis J. O'Reilly[1], Michael Bohlke-Schneider[1], Adam Belsom[2], Therese Dau[2], Colin W. Combe[2], Martin Graham[2], Markus R. Eisele[5], Wolfgang Baumeister[5], Christian Speck[3,4], Juri Rappsilber[1,2,*]

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

[2]Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

[3]MRC London Institute of Medical Sciences (LMS), London, UK

[4]DNA Replication Group, Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, UK

[5]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[#]Marta L. Mendes and Lutz Fischer contributed equally to this work

*Corresponding author: juri.rappsilber@tu-berlin.de

We present a concise workflow to enhance the mass spectrometric detection of cross-linked peptides by introducing sequential digestion and the cross-link identification software Xi. Sequential digestion enhances peptide detection by shortening long tryptic peptides while avoiding over-digestion. We demonstrate our simple 12-fraction protocol for cross-linked multi-protein complexes and cell lysates, quantitative analysis, and high-density cross-linking, without requiring specific cross-linker features. This overall approach reveals dynamic protein-protein interaction sites, which are accessible, have fundamental functional relevance and are therefore ideally suited for the development of small molecule inhibitors.

Cross-linking/mass spectrometry (CLMS) has become a standard tool for the topological analysis of multi-protein complexes[1] and has begun delivering high-density information on protein structures[2], insights into structural changes[3] and the wiring of interaction networks *in situ*[4]. The technological development currently focuses on enrichment strategies for cross-linked peptides and mass spectrometric data acquisition[5–7], including newly designed cross-linkers[8]. MS2-

cleavable cross-linkers in particular, have celebrated recent successes for the analysis of protein complexes[9] or complex mixtures[10,11]. Functionalized cross-linkers typically form part of an entire workflow including analysis software tailored to the data formats obtained, such as XLinkX[10] but, importantly, there are also flexible general software that are not paired with distinct cross-linker properties, such as pLink[12], StavroX[13] and Kojak[14].

The focus on bespoke cross-linkers has left general steps of sample preparation, such as protein digestion, with less attention. Tryptic digestion generates cross-linked peptides of considerable size; a quality that has been exploited with their enrichment by SEC[5], but one that poses as a potential problem regarding their detection. Replacing trypsin with proteases such as GluC, AspN and chymotrypsin does not change this[15,16]. We reasoned that sequential digestion could reduce the size of large tryptic peptides and offer access to sequence space that otherwise would remain undetected. We therefore followed trypsin digestion with subsequent digestion by alternative proteases and developed Xi, a database search engine, allowing the search of multiple datasets resulting from the application of our protocol. This novel approach expands the detectable structure space in proteins, allowing it to capture dynamic regions in protein complexes that are mechanistically important and druggable, however that hitherto have remained undisclosed by cryo-EM.

We first tested this workflow on a standard mix of seven Bis[sulfosuccinimidyl] suberate (BS^3) cross-linked proteins (Catalase, Myoglobin, Cytochrome C, Lysozyme, Creatine Kinase, HSA and Conalbumin). Importantly, their structures are known and hence offer an independent assessment of false identifications. Four digestion conditions, each giving three SEC fractions, resulted in a total of 12 acquisitions, which is the protocol applied to all subsequent analyses presented here (Fig. 1a). We compared the results to samples digested with trypsin alone and to a protocol that used the four proteases individually, each analysed with 12 injections to reach comparable analytical effort. Sequential digestion produced the best results, increasing the

number of residue pairs by 1.5-fold without compromising the agreement of CLMS with the crystal structures (Supplementary Fig. 1a, e, f, Supplementary Data 1). Before assessing if this improvement translated into a gain of information in a number of biological applications we investigated the origin of the added data.

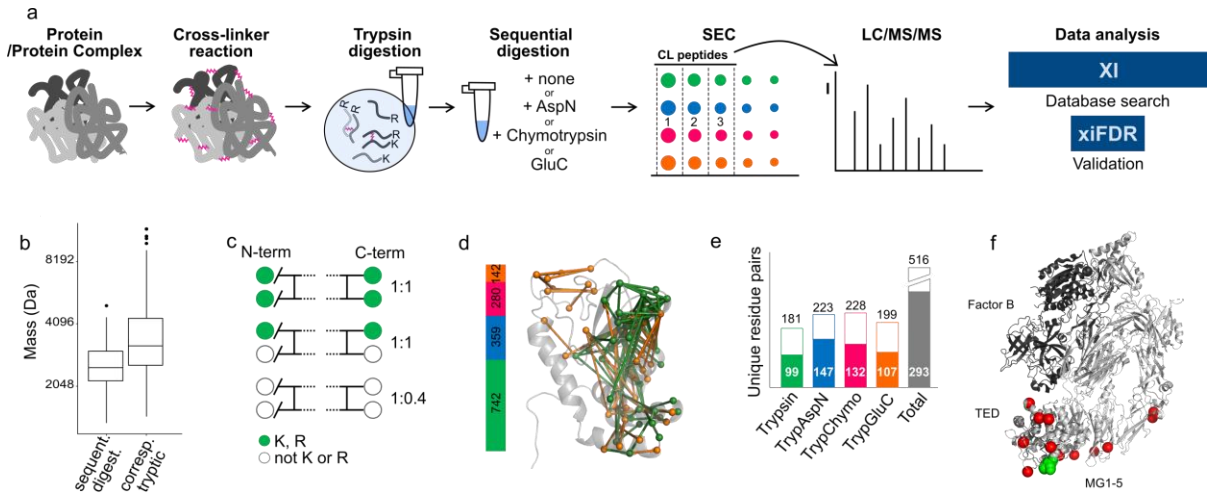

Figure 1: *Sequential digestion strategy and its impact on different sample complexities, cross-linker chemistry and quantitation. (a) Sequential digestion workflow. Proteins or protein complexes are cross-linked and digested with trypsin. After splitting the sample into four aliquots, one is left alone while the others are sequentially digested with either AspN, chymotrypsin or GluC. Samples are enriched by SEC and the three high-MW fractions are analysed by LC-MS, submitted to Xi search and xiFDR for data validation. (b) Cross-linked peptides obtained by sequential digestion are smaller than their corresponding tryptic peptides. (c) Sequential digested cross-linked peptides show a bias towards having C-termini that end in K or R, which is not the case for the N-termini, showing that sequential digestion delivers smaller tryptic peptides easier to identify by LC-MS/MS. (d) Sequential digestion workflow applied to high-density CLMS of CASP12 target, a 193 residue section of UGGT, (PDB|5NV4) with unique residues gained (bar) by adding AspN (blue), chymotrypsin (pink) and GluC (orange) to trypsin (green). Additional residue pairs identified upon sequential digestion (orange) cover regions of the protein structure not observed with trypsin (green). (e) Sequential digestion applied to QCLMS of C3b dimerization. Half of the identified residue pairs*

per fraction were quantified. (f) C3b (light grey) bound to factor B (dark grey) (PDB|2XWJ) with position of thioester (green) and position of C3b dimer-specific links (red).

Indeed, sequential digestion led to smaller peptides than trypsin alone (Fig. 1b). Notably, we observed a bias towards having tryptic C-termini but not N-termini (Fig. 1c). This identification bias is consistent with better fragmentation behaviour of peptides with basic C-terminus[17] and shows the importance of trypsin as part of the protocol. Importantly, sequential digestion does not lead to the very short peptides one might expect from the action of two proteases (Fig. 1b). This agrees with reports that serine proteases lose efficiency as peptides shorten[18-20]. Consequently, sequential digestion brings large tryptic peptides into a more detectable size range without overly shortening peptides, while maintaining the presence of C-terminal basic residues.

We then tested the sequential digestion approach on samples of increasing complexity ranging from single proteins, UGGT and C3b, to the OCCM DNA replication complex (1.1 MDa), the 26S proteasome (2.5 MDa) and high-molecular weight fractions of human cytosol. A quantitative experiment was performed to assess the efficiency of sequential digestion combined with the QCLMS workflow[21]. Additionally, we tested the approach using two different cross-linkers, the homobifunctional cross-linker BS^3 and the heterobifunctional, photoactivatable cross-linker sulfosuccinimidyl 4,4´-azipentanoate (SDA).

UGGT was one of the data-assisted *de novo* folding targets of CASP12 for which we contributed data in the form of 433 unique residue pairs obtained at a 5% FDR (http://predictioncenter.org/download_area/CASP12/extra_experiments/) (Supplementary Fig. 2a) using SDA as cross-linker and 26 LC-MS runs[22]. Using sequential digestion, we now identified 1523 unique residue pairs in only 12 runs (Fig. 1d, Supplementary Data 1). With 5% long-distance links (> 20 Å) when mapped onto the structure released by CASP organizers (Fig. 1d, Supplementary Fig. 2c) the 3-fold increase in observed links comes at uncompromised reliability. Consequently, the sequential digestion protocol improves high-density CLMS by a clear increase

in the number of residue pairs while simultaneously reducing the analytical effort needed to detect these.

We next combined quantitative CLMS (QCLMS)[3,23,24] with our workflow (Supplementary Fig. 3a) to investigate the dimerization of C3b. Thioester-mediated dimerization of C3b is a key process of the human complement response enhancing the efficiency of C5 convertase formation which ultimately leads to clearance of pathogens from human blood[25-27]. However, the structure of this dimer is currently unclear. The reactive thioester could result in a random orientation of the two C3b molecules in a dimer. Alternatively, auxiliary factors or self-organisation properties of C3b could mediate a preferred orientation. We here investigate C3b alone and find it to form dimers in the absence of active thioester and auxiliary proteins. We quantified 293 unique cross-links, about three times more than with trypsin alone (99) (Fig 1e, Supplementary Data 1) which lends robust support to a bottom-to-bottom orientation (Fig1f). This suggests non-covalent interactions between C3b molecules lead to a preferred dimer orientation which implies that a thioester bridged dimer would follow this arrangement. Non-covalent interactions thus self-organise C3b into a productive dimer as this arrangement is compatible with the subsequent molecular events of the complement cascade by allowing unhindered binding of factor B at the top of C3b.

Turning our attention to protein complexes we investigated the OCCM complex, a helicase-loading intermediate formed during the initiation of DNA replication. Recently a 3.9-Å structure of *Saccharomyces cerevisiae* OCCM on DNA was obtained by cryo-electron microscopy (cryo-EM)[28] supported by CLMS (Supplementary Fig. 4a, Supplementary Data 1). We identified 682 residues pairs from the same sample analysed before, with large contribution from sequential digestion (Fig 2a, Supplementary Fig 4b). Interactions observed now include known Cdt1-Mcm2 and Mcm6 but also Mcm2-Orc5 interaction (Mcm2-850-Orc5-369). These led us to delete the C-terminal 20 aa of Mcm2 (848-868) (Fig. 2C lane 5) and analyse its biological relevance in a well-

established *in vitro* helicase loading assay, which recapitulates the *in vivo* process[29]. The deletion mutant did not affect ORC, Cdc6, Cdt1 and origin DNA dependent complex assembly under low salt conditions (Fig. 2b, lanes 6 and 7), but severely impaired the formation of the high salt stable double-hexamer (Fig. 2b, lanes 8 and 9), the final product of the helicase loading reaction. This is an exciting result, as it highlights a novel and essential role for Mcm2 C-terminus in a late step of MCM2-7 double-hexamer formation, a process that is only poorly understood. Moreover, the CLMS data show that the Mcm2 C-terminus is involved in a network of interactions with flexible domains of Orc6, Orc2 and Mcm5, which could represent an ideal target for development of inhibitors with potential as anti-cancer therapy[30], as dynamic interactions have improved druggability characteristics over stable protein interactions[31,32]. Indeed, expressing Mcm2-7ΔC2 causes dominant lethality (Fig. 2c). The ability of CLMS data to complete the cryo-EM structure of the OCCM complex by dynamic contacts proved here essential. Note that 15% of our residue pairs falling into the published OCCM structure were long distance (> 30 Å, Supplementary Fig. 4c). This indicates that CLMS unveils dynamic aspects of protein complex topology also in regions of the structure accessible to cryo-EM as will become even more evident in our proteasome analysis.

We next analysed an affinity-purified 26S proteasome sample, containing more than 600 proteins (Supplementary Data 2). The results of our workflow compare favourably with the largest analysis reported on this complex to date[9] in terms of numbers (n=1644 versus 447 unique residue pairs in the proteasome at 5% FDR) (Fig. 2d,e, Supplementary Fig. 5a, Supplementary Data 1) and in terms of agreeing with the structure of the individual subunits (6% versus 26% long-distance links (> 30 Å)) (Fig. 2d,e). Links between proteins (n=602) reveal a large amount of topological variability in the proteasome, with 30% (n=179) being not covered by current cryo-EM based models and thus extending our awareness of the proteasome structure to more dynamic regions. Long distance links (n= 191 between and 85 within proteins) are mainly distributed in the base of the proteasome, where ATP-binding and hydrolysis leads to a large conformational variety

(Fig. 2f). Indeed, some of these links (n=78) not matching to one structure of the proteasome mapped well to alternative conformational states stabilised by ATP analoga[33,34]. State-specific cross-links were found predominantly in the AAA-ATPase dependent heterohexameric ring (Supplementary Fig. 5) indicating rearrangement of Rpn5 relative to Rpt4 (Fig. 2g). In the s2 state our data support Rpn1 being translated and rotated to be positioned closer to the AAA-ATPase (Fig. 2g). Cross-links therefore support in solution the cryoEM-based model of substrate transfer to the mouth of the AAA-ATPase heterohexameric ring[33] and point towards the existence of additional conformational states that remain to be defined to fully understand the complex's function and that may offer as conformer-specific interactions prime intervention points for small molecule inhibitors.

To probe our 12-fraction protocol in large-scale CLMS we analysed seven high-molecular weight fractions of human cytosol. We identified 3572 unique residue pairs (5% FDR, 528 proteins, Supplementary Fig. 6a, Supplementary Data 1). This is in line with recent studies reporting 1684 and 3045 unique residue pairs, respectively, albeit using a cleavable cross-linker[7,10]. Interestingly, the overlap between the published data and ours is strikingly low (Fig. 2h) suggesting the approaches are largely complementary.

Our protein-protein interaction network included previously observed complexes like the Mcm2-7 complex, the 26S proteasome, the ribosome, the COPI complex, the TRiC-CCT complex and the HS90B-CDC37-Cdk4 complex (Fig. 2i and Supplementary Fig. 6b-g, 7). For the 26S proteasome we were able to distinguish between different states defining flexibility in the AAA-ATPase ring (Supplementary Fig. 6g-k)[35]. This indicates the ability of our protocol to unveil dynamic interactions in mixtures nearing the native environment complexity of proteins.

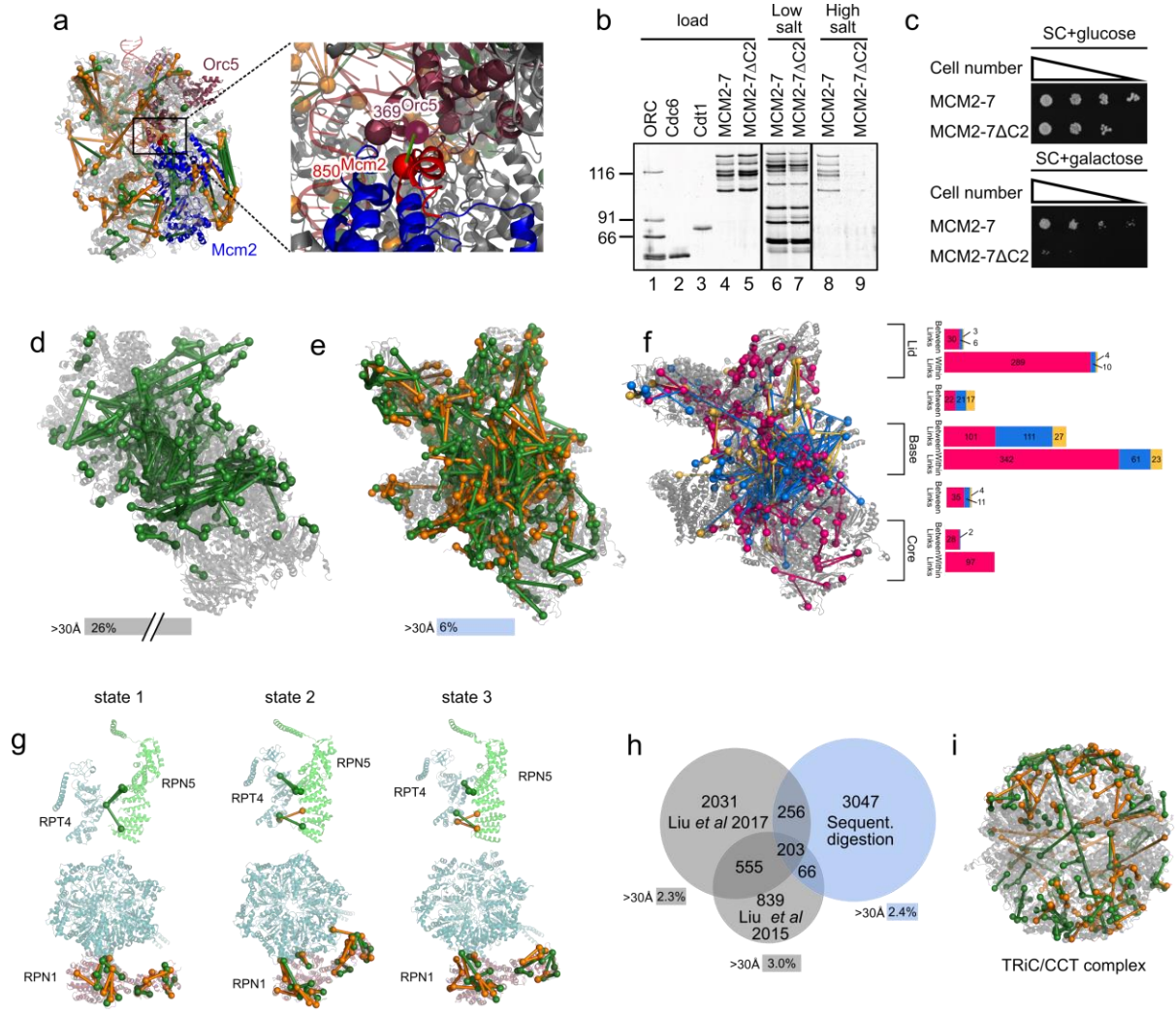

Figure 2: *Sequential digestion of the affinity-purified complexes OCCM (*S. cer.*) and 26S Proteasome (*S. cer.*) and the cytosol of human K562 cells. Residue pairs observed in tryptic (green) and non-tryptic (orange) peptides. (a) Unique residue pairs mapped to the OCCM complex (PDB|5udb) and the key link Mcm2 – Orc5 (Mcm2-850-Orc5-369). (b) The in vitro helicase loading assay demonstrates that an Mcm2 C-terminal deletion mutant supports complex assembly (lanes 6 and 7) and blocks formation of the final helicase loading product (lanes 8 and 9). (c) Overexpression analysis of Mcm2-7ΔC2 shows that this mutant causes dominant lethality, indicating that the C-terminus of Mcm2 is essential in cell survival. (d) Unique residue pairs obtained by Wang* et al. *for the human 26S Proteasome (PDB|5GJR). (e) Unique residue pairs obtained by sequential digestion obtained for the S. cerevisiae 26S proteasome (PDB|4CR2). Sequential digestion returned the highest number of residue pairs so far identified by CLMS for the 26S proteasome. (f) Long distance (blue) and within distance (pink) between residue pairs were mapped into one of the states*

of the proteasome (4cr2) showing the accumulation of those into the base of the complex. Residue pairs satisfying other states are represented in yellow. The bar plot shows the distribution of all residue pairs in the complex showing that long distance links locate mainly in the base. (g) Unique residue pairs were mapped into the 3 states described by Unverdorben et al. showing that different conformations can be distinguished by using sequential digestion. (h) Comparison of sequential digestion for complex mixtures with previous studies. Sequential digestion returns a higher number of residue pairs with low overlap to published datasets, showing the complementarity of the different approaches. Over-length links were determined only within proteins and are comparable for all datasets. (i) Residue pairs for the TRiC/CCT complex were mapped into the crystal structure and support the rearrangement of the complex reported by Leitner et al.[1](PDB|4V94).

To analyse the mass spectrometric data of these and other studies[7,9,10] we developed our database search software Xi (Fig 3a, Supplementary Fig. 8, http://xi3.bio.ed.ac.uk/downloads/xiSearch/). The algorithm of Xi has been described conceptually before[36]. It follows an approach that computationally unlinks cross-linked peptides and by doing so circumvents the n2 database problem of cross-linking. Like StavroX[13] and Kojak[14], Xi allows to search any cross-link and protease specificity, thus Xi's performance was assessed against these two alternatives. StavroX became non-responsive and was not pursued further. Kojak was paired with Percolator[37] and Xi with xiFDR[38] to maximise results and control the error rate to 5% FDR. Kojak/Percolator identified 1529 unique residue pairs and Xi/xiFDR 2098 (+40%) (Fig 3b, Supplementary Data 1). Structure mismatch for self-links was 12% for Kojak and 6% for Xi. In summary, Xi performed very favourably compared to the universal software for CLMS.

Our integrated workflow utilises standard cross-linkers, without special chemistries to assist analysis. While recent large-scale studies have successfully used cleavable cross-linkers[10,39], our work uses standard cross-linkers at no obvious disadvantage. MS-cleavable cross-linkers have yet to be combined with high-density cross-linking and are likely incompatible with cross-linking by non-canonical amino acids[40] motivating efforts in keeping cross-link

chemistry and analysis workflows separate. Our protocol supports this drive and provides a concise, universal protocol to increase data density and ease of use for CLMS in diverse applications that include the detection of dynamic protein interaction regions and topologies which are notoriously difficult to detect using conventional structural biology methods yet are prime therapeutic intervention points.

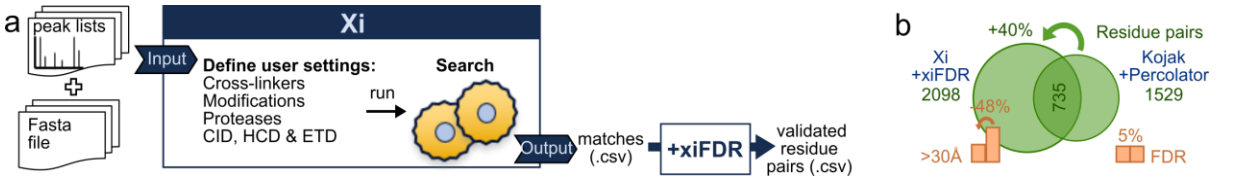

Figure 3: *Xi search engine. (a) Xi is an open source search engine that takes a peak list as input. Users can define any type of cross-linker, modification, digestion, and fragmentation method. The output is a list of matches in .csv format. We use xiFDR to validate our results (b) Xi and Kojak comparison. The same dataset of the 26S Proteasome was searched in both softwares and data validation was performed by xiFDR for Xi and by Percolator for Kojak. At 5% FDR, Xi and xiFDR, delivered 40% more links with better quality (fraction of long distance links within proteins returned by Kojak: 11.6%; Xi: 6%).*

Acknowledgements

We thank Swantje Lenz for the technical support and helpful discussions and Petra Ryl for sample preparation support. We thank CASP12 organizers for supplying UGGT sample and Alberto Riera for contributing OCCM sample. M.L.M. was supported by The International Post-Doc Initiative - IPODI, co-funded by the European Union. This work was supported by the Einstein Foundation, the DFG [RA 2365/4-1], and the Wellcome Trust through a Senior Research Fellowship to J.R. [103139], an Investigator Award to C.S. [107903/Z/15/Z] and a multi-user equipment grant to J.R.

[108504], the Biotechnology and Biological Sciences Research Council UK to C.S. [BB/N000323/1] and the Medical Research Council UK to C.S. [MC_U120085811]. The Wellcome Centre for Cell Biology is supported by core funding from the Wellcome Trust [203149].

Author contribution

M.L.M, L.F. and J.R. developed the study; M.L.M, Z.A.C., F.O'R., A.B, M.B and M.R.E. performed sample preparation; M.L.M. performed LC-MS analysis; M.L.M., Z.A.C., F.O'R., A.B., T.D., M.B., C.S and M.B.S performed data analysis; L.F. developed Xi; C.W.C and M.G. provided critical support in data analysis; M.B., C.S., M.R.E. and W.B. contributed with data interpretation; M.L.M., F.O'R, M.B.S, L.F. and J.R. wrote the article with critical input from all authors.

Competing interests

The authors declare no competing financial interests.

References

1. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **41,** 20–32 (2016).
2. Belsom, A., Schneider, M., Fischer, L., Brock, O. & Rappsilber, J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Mol. Cell. Proteomics* **15,** 1105–1116 (2016).
3. Chen, Z. A. *et al.* Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling. *Mol. Cell. Proteomics* **15,** 2730–2743 (2016).

4. Schweppe, D. K. *et al.* Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 1732–1737 (2017).
5. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.* **9,** 120–137 (2013).
6. Kolbowski, L., Mendes, M. L. & Rappsilber, J. Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. *Anal. Chem.* **89,** 5311–5318 (2017).
7. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8,** 15473 (2017).
8. Kao, A. *et al.* Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics* **10,** M110.002212 (2011).
9. Wang, X. *et al.* Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Mol. Cell. Proteomics* (2017). doi:10.1074/mcp.M116.065326
10. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12,** 1179–1184 (2015).
11. Chavez, J. D., Weisbrod, C. R., Zheng, C., Eng, J. K. & Bruce, J. E. Protein interactions, post-translational modifications and topologies in human cells. *Mol. Cell. Proteomics* **12,** 1451–1467 (2013).
12. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9,** 904–906 (2012).
13. Götze, M. *et al.* StavroX--a software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **23,** 76–87 (2012).
14. Hoopmann, M. R. *et al.* Kojak: efficient analysis of chemically cross-linked protein complexes. *J. Proteome Res.* **14,** 2190–2198 (2015).

15. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteomics* **11,** M111.014126 (2012).
16. Schilling, B., Row, R. H., Gibson, B. W., Guo, X. & Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14,** 834–850 (2003).
17. Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3,** 608–614 (2004).
18. Wenzel, H. R. & Tschesche, H. Cleavage of peptide-4-nitroanilide substrates with varying chain length by human leukocyte elastase. *Hoppe Seylers Z. Physiol. Chem.* **362,** 829–831 (1981).
19. Fiedler, F. Effects of secondary interactions on the kinetics of peptide and peptide ester hydrolysis by tissue kallikrein and trypsin. *Eur. J. Biochem.* **163,** 303–312 (1987).
20. Hill, C. R. & Tomalin, G. The kinetics of hydrolysis of some extended N-aminoacyl-l-phenylalanine methyl esters by bovine chymotrypsin Aα Evidence for enzyme subsite S5. *Biochimica et Biophysica Acta (BBA) - Enzymology* **660,** 65–72 (1981).
21. Chen, Z. *et al.* Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. *Wellcome Open Res* **1,** 5 (2016).
22. Ogorzalek, T. L. *et al.* Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. *Proteins* (2018). doi:10.1002/prot.25452
23. Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat. Commun.* **4,** 1985 (2013).
24. Tomko, R. J., Jr *et al.* A Single α Helix Drives Extensive Remodeling of the Proteasome Lid and Completion of Regulatory Particle Assembly. *Cell* **163,** 432–444 (2015).
25. Hong, K. *et al.* Reconstitution of C5 convertase of the alternative complement pathway with isolated C3b dimer and factors B and D. *J. Immunol.* **146,** 1868–1873 (1991).

26. Pangburn, M. K. & Rawal, N. Structure and function of complement C5 convertase enzymes. *Biochem. Soc. Trans.* **30,** 1006–1010 (2002).

27. Rawal, N. & Pangburn, M. K. Structure/function of C5 convertases of complement. *Int. Immunopharmacol.* **1,** 415–422 (2001).

28. Yuan, Z. *et al.* Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat. Struct. Mol. Biol.* **24,** 316–324 (2017).

29. Evrin, C. *et al.* A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20240–20245 (2009).

30. Gardner, N. J. *et al.* The High-Affinity Interaction between ORC and DNA that Is Required for Replication Licensing Is Inhibited by 2-Arylquinolin-4-Amines. *Cell Chem Biol* **24,** 981–992.e4 (2017).

31. Ulucan, O., Eyrisch, S. & Helms, V. Druggability of dynamic protein-protein interfaces. *Curr. Pharm. Des.* **18,** 4599–4606 (2012).

32. Jubb, H., Blundell, T. L. & Ascher, D. B. Flexibility and small pockets at protein-protein interfaces: New insights into druggability. *Prog. Biophys. Mol. Biol.* **119,** 2–9 (2015).

33. Unverdorben, P. *et al.* Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 5544–5549 (2014).

34. Wehmer, M. *et al.* Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 1305–1310 (2017).

35. Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 12991–12996 (2016).

36. Giese, S. H., Fischer, L. & Rappsilber, J. A Study into the Collision-induced Dissociation (CID) Behavior of Cross-Linked Peptides. *Mol. Cell. Proteomics* **15,** 1094–1104 (2016).

37. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4,** 923–925 (2007).

38. Fischer, L. & Rappsilber, J. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal. Chem.* **89,** 3829–3833 (2017).

39. Hage, C., Falvo, F., Schäfer, M. & Sinz, A. Novel Concepts of MS-Cleavable Cross-linkers for Improved Peptide Structure Analysis. *J. Am. Soc. Mass Spectrom.* **28,** 2022–2038 (2017).

40. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2,** 261–267 (2005).

Materials and Methods

Reagents

Catalase, Myoglobin, Cytochrome C, Lysozyme, human Serum Albumin (HSA) and Conalbumin were purchased from Sigma. Creatine Kinase was purchased from Roche. C3b was purchased from Complement Technology, Inc. UGGT was kindly provided by the *The Critical Assessment of* protein *Structure Prediction* experiment organizers. Bis[sulsosuccinimidyl]suberate (BS^3), Disuccinimidyl suberate (DSS), sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA) and Trypsin were purchased from Thermo Scientific. GluC, Chymotrypsin and AspN, were purchased from Promega. All media, supplements and phosphate buffered saline (PBS) for cell culture were purchased from PAA Laboratories.

Sample preparation, cross-linking and individual digestions

Catalase, Myoglobin, Cytochrome C, Lysozyme, Creatine Kinase, HSA and Conalbumin were resuspended in BS3 cross-linking buffer (20 mM HEPES, 20 mM NaCl, 5 mM $MgCl_2$, pH 7.8) to a final concentration of 1 mg/ml. Cross-linker was added to a 1:1 (w/w) protein to cross-linker ratio and samples incubated for 2 h on ice. Cross-linking reaction was quenched with excess ABC for 1 h at room temperature (RT). Cross-linked samples were loaded on a NuPAGE™ 4-12% Bis-Tris protein gel to isolate the monomeric band that was then in-gel digested with trypsin[41]. After peptide extraction from the gel, samples were mixed in a 1:1 weight ratio to a final amount of 200 µg. Sample was divided in four parts and desalted using C18 StageTips[42]. For digestion with trypsin and individual digestions with AspN, chymotrypsin and GluC the procedure was the same as above. Proteases were added as follows: GluC and chymotrypsin were added to a protease-to-substrate (w/w) ratio of 1:50 and incubated overnight (ON) at 37°C and RT, respectively; AspN

was added to a protease-to-substrate ratio (w/w) of 1:100 and incubated ON at 37°C. After digestion samples were fractionated by SEC and analysed by LC-MS/MS.

For the OCCM complex pUC19-ARS1 beads were used to assemble the OCCM complex as described elsewhere[43]. Cross-linking was performed on beads. BS3 was added to 200 µg of the OCCM complex to a 1:8100 protein to cross-linker molar ratio. The sample was incubated for 2 h on ice and the cross-linking reaction was quenched with excess ABC for 1h at RT. The sample was transferred into 8 M Urea, reduced with Dithiothreitol (DTT), alkylated with iodoacetamide (IAA) and diluted with ABC 50 mM to a final concentration of 2 M Urea. Trypsin was added to a protease-to-substrate ratio of 1:50 and the sample was incubated ON at 37°C.

The 26S Proteasome was isolated from *Saccharomyces cerevisiae* by affinity purification using the 3x FLAG-tagged subunit Rpn11 as described elsewhere[44]. For the cross-linking, the 26S Proteasome buffer was exchange to BS3 cross-linking buffer using 30 kDa molecular weight cutoff (MWCO) filters (Milipore). 200 µg of the 26S Proteasome were cross-linked with BS3. BS3 was added to a 1:1 (w/w) protein to cross-linker ratio. Samples were incubated for 2 h on ice and the cross-linking reaction was quenched with excess ABC for 1 h at room temperature (RT). The sample was dried using a vacuum concentrator and resuspended in 6 M Urea/2 M Thiourea for subsequent *in-solution* digestion. Sample was reduced with 2.5 mM DTT for 15 min at 50°C, then alkylated with 5 mM IAA at RT in the dark and diluted with ABC 50 mM to a final concentration of 1 M. Trypsin was added at an enzyme-to-substrate mass ratio of 1:50 and the sample was incubated ON at 37°C. Reaction was stopped with 10% (v/v) TFA, and the samples divided into four parts and desalted using StageTips.

K562 cells (DSMZ, Cat# ACC-10, negatively tested for mycoplasma) were grown in T175 flasks at 37°C in humidified 5% (v/v) CO_2 incubators in RPMI 1640 media supplemented with 10% (v/v) fetal bovine serum (FBS) + 2 mM glutamine. $3x10^8$ cells were harvested by centrifugation (180xg) and washed 3 times with ice cold PBS. Cells were lysed in lysis buffer (100 mM HEPES

pH 7.2, 100 mM KCl, 20 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, 10 µg/ml DNAse1, Complete EDTA-free protease inhibitor cocktail, 1 mM PMSF) using a douncer at 4°C. Lysate was cleared of debris by centrifugation at 100,000 x g for 45 min. Native protein complexes were further concentrated by spin filtration using a 100,000 Da cut-off Amicon Ultra centrifugal filter unit (approx. 30 min). For protein co-elution analysis, 100 µl of concentrated lysate (approximately 30 mg/ml) was separated by a Biosep SEC-S4000 (7.8 x 600) size exclusion column on an Åkta Purifier (HPLC) system running at 0.25 ml/min 100 mM HEPES pH 7.2, 100 mM KCl, 20 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA, 10% glycerol. 500 µl fractions were collected. Protein fractions were cross-linked at <1mg/ml (quantified by Bradford) and with 2:1 w/w ratio of BS3 to protein. Samples were *in-solution* digested with trypsin as described above.

C3b monomer and dimer were labelled as described elsewhere[20] and in gel digested with trypsin.

UGGT was cross-linked using Sulfo-SDA using eight different protein to cross-linker ratios (1:0.13, 1:0.19, 1:0.25, 1:0.38, 1:0.5, 1:0.75, 1:1 and 1:1.5 (w/w)). Cross-linking was carried out in two-stages: firstly sulfo-SDA, dissolved in SDA cross-linking buffer (25 µl, 20 mM HEPES-OH, 20 mM NaCl, 5 mM $MgCl_2$, pH 7.8), was added to target protein (25 µg, 1 µg/µl) and left to react in the dark for 50 minutes at room temperature. The diazirine group was then photo-activated using UV irradiation, at 365 nm, from a UVP CL-1000 UV Crosslinker (UVP Inc.). Samples were spread onto the inside of Eppendorf tube lids by pipetting (covering the entire surface of the inner lid), placed on ice at a distance of 5 cm from the tubes and irradiated for 20 minutes. The reaction mixtures from the eight conditions corresponding to each experiment were combined and quenched with the addition of saturated ABC (13 µl). Sample was dried in a vacuum concentrator and 200 µg of each protein were *in-solution* digested as described above.

Sequential digestion

After trypsin digestion reactions were stopped with 10% (v/v) TFA, divided into four parts and desalted using StageTips. Eluted peptides were dried and while one part was fractionated by SEC,the remaining three were sequential digested.

Each one of the samples for sequential digestion was resuspended in Ammonium Bicarbonate (ABC) 50 mM and sequential digested as follows: for the sequential digestion with GluC and chymotrypsin, proteases were added to an enzyme-to-substrate ratio of 1:50 and incubated ON at 37°C and RT, respectively. For the sequential digestion with AspN, the enzyme was added at an enzyme-to-substrate ratio of 1:100 and incubated ON at 37°C. After sequential digestion samples were acidified using 10% TFA and the sample volume was reduced to 50 µl by evaporation using a vacuum concentrator. Samples were fractionated by SEC.

Fractionation of peptides by Size Exclusion Chromatography (SEC)

Size exclusion chromatography of cross-linked peptides was performed as described elsewhere[5]. 50 µg of peptides were fractionated in a Shimadzu HPLC system using a Superdex Peptide 3.2/300 (GE Healthcare) at a flow rate of 50 µl/min using SEC buffer (30% (v/v) ACN, 0.1% (v/v) TFA) as mobile phase. Separation was monitored by UV absorption at 215 and 280 nm. Fractions were collected every two minutes over one column volume. The three high-MW fractions were dried, resuspended in 0.1% (v/v) TFA and analysed by LC-MS/MS.

In vitro pre-RC assay

The *in vitro* pre-RC assay was performed as described[45,46]. Briefly, ORC (40 nM), Cdc6 (80 nM), Cdt1 (40 nM), MCM2-7 (40 nM) or MCM2-7 ΔC2 (40 nM) were incubated with 6 nM of DNA containing the ARS1 sequence in 50 µl of pre-RC buffer (50 mM HEPES pH 7.5, 100 mM potassium glutamate, 10 mM magnesium acetate, 50 µM zinc acetate, 3 mM ATP, 5 mM DTT,

50 µM EDTA, 0.1% triton-X, 5% glycerol). After 20 minutes at 24 °C, the reactions were washed three times with low salt buffer (pre-RC buffer) or high salt buffer (pre-RC buffer plus 300 mM NaCl) before digestion with 1 U of DNaseI in pre-RC buffer plus $CaCl_2$.

Yeast lethality assay

Yeast strain AS499 (MATa. bar1Δ, leu2-3,-112, ura3-52, his3-Δ200, trp1-Δ-63, ade2-1 lys2-801, pep4) was transformed with pESC-LEU-MCM2-MCM7, pESC-TRP-MCM6-MCM4 and pESC-URA-HA-MCM3-MCM5 (wild type, YC119) or pESC-LEU-MCM2 (Δ848-868)-MCM7 (MCM2-7ΔC2, YC388). Yeast strains YC119 and YC388 were plated on a dropout synthetic complete (SC) medium and incubated at 30 °C for 48 hours. Cells were grown in suspension to 10^7 cells/ml. Three microliters of a five-fold serial dilutions were spotted onto selective plates containing either 2% galactose or glucose. Plates were incubated at 30 °C for 3-5 days.

LC-MS/MS

Samples were analysed using a Thermo Scientific Dionex Ultimate 3000 RSLCnano system coupled to a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with an EASY-Spray source. Mobile phase A consisted in 0.1% (v/v) FA in water and mobile phase B consisted of 80% (v/v) ACN and 0.1% (v/v) FA in water. Peptides were loaded into a 50 cm EASY-Spray column operated at a temperature of 45°C at a flow rate of 300 nl/min and separated at 300 nl/min using the following gradient: 2% mobile phase B (0-11 min); 2-40% mobile phase B (11-150 min); 40-95% mobile phase B (150-161 min); 95% mobile phase B (161-166 min); 95-2% mobile phase B (166-185 min).

MS data were acquired using a “high-high” acquisition method using the Orbitrap to detect both MS and MS/MS scans. The instrument was operated in a data dependent mode with a cycle time of 3 s. MS1 scans were acquired at a resolution of 120,000 using a scan range from 300 to

1700 m/z and AGC target of $2.5x10^5$ with a maximum injection time of 50 ms. The Monoisotopic Peak Determination (MIPS) was activated and only precursors with charge states between 3 and 8 with an intensity threshold higher than $5.0x10^4$, were selected for fragmentation. Selected precursors were fragmented by HCD using an HCD Collision Energy of 30%. MS2 spectra were acquired at a resolution of 15,000 and AGC of 1.0x104 with a maximum injection time of 35 ms. Dynamic exclusion was set to 60 s after 1 count.

Data analysis

Thermo raw data were pre-processed using MaxQuant (v 1.5.7.4) to extract the peak list files (APL format). Partial processing in MaxQuant was performed until step 5 with the parameters set to default with exception of the "FTMS top peaks per Da interval" which was set to 100 and no FTMS de-isotoping was allowed. apl files were uploaded to Xi for identification of cross-linked peptides (Xi software is available from http://xi3.bio.ed.ac.uk/downloads/xiSearch/ and the code is available from https://github.com/Rappsilber-Laboratory/XiSearch). For the cross-linking search the parameters used were as follows: MS accuracy, 6 ppm; MS/MS accuracy, 20 ppm; enzyme, trypsin, trypsin+AspN, trypsin+chymotrypsin or trypsin+Gluc depending on the sample digestion conditions; missed cleavages allowed, 4. For the BS3 cross-linked samples, carbamidomethylation of cysteines was set as a fixed modification and oxidation of methionines and the cross-linker alone (mass modification: 109.0396 Da), hydrolyzed (BS3-OH: 156.0786 Da) or amidated (BS3-NH2: 155.0964 Da) were set as variable modifications. Reaction specificity for BS3 was assumed to be with Lysine, Serine, Threonine and Tyrosine or the protein N-terminus. For SDA cross-linked samples, carbamidomethylation of cysteines, oxidation of methionines and the cross-linker alone (mass modification: 109.0396 Da), hydrolyzed (SDA-OH: 100.0524 Da) or cross-linker loop (SDA-loop: SDA cross-link within a peptide that is also cross-linked to a separate peptide, 82.0419 Da) were set as variable modifications. Reaction specificity for SDA was

assumed to be with Lysine, Serine, Threonine and Tyrosine or the protein N-terminus on one end of the spacer and with all residues on the other end. Apl files were searched against the following databases: for the Standard Protein Mix a database was built containing the sequences corresponding to the crystal structures of the standard proteins used in the mix (PDB accession: 3j7u, 5d5r, 3nbs, 1dpx, 2crk, 1ao6 and 1ova); for C3b we used the fasta corresponding to the uniprot accession P01024; for the OCCM complex a database containing all the 14 OCCM subunits was used; for the 26S Proteasome a linear search of the sample was first performed using a complete *S. cerevisiae* database and the 1% most intense proteins accordingly to the iBAQ value were selected to build the database for the cross-linking search; for each one of the cytosolic SEC fractions the procedure was as for the 26S Proteasome, using as database for the linear searched the complete human database; for the UGGT, the uniprot sequence of the protein was used (G0SB58). FDR was estimated on a 5% residue level, including only unique PSMs and boosting results, using xiFDR. Necessary alignments were made using the score matrix BLOSUM62.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[47] with the dataset identifier <PXD008550>.

Additional References

41. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1,** 2856–2860 (2006).
42. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670 (2003).

43. Evrin, C. *et al.* A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20240–20245 (2009).

44. Sakata, E. *et al.* The catalytic activity of Ubp6 enhances maturation of the proteasomal regulatory particle. *Mol. Cell* **42,** 637–649 (2011).

45. Evrin C, et al. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci USA **106**, 20240–20245, (2009).

46. Fernández-Cid A, et al. An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. Mol Cell **50**, 577–588 (2013).

47. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA,Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolomé S, Apweiler R,Omenn GS, Martens L, Jones AR, Hermjakob H. ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination. *Nat. Biotechecnol*. **32**, 223-226 (2014).