



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## An integrated workflow for crosslinking mass spectrometry

### Citation for published version:

Mendes, ML, Fischer, L, Chen, ZA, Barbon, M, O'Reilly, FJ, Giese, SH, Bohlke-Schneider, M, Belsom, A, Dau, T, Combe, CW, Graham, M, Eisele, MR, Baumeister, W, Speck, C & Rappsilber, J 2019, 'An integrated workflow for crosslinking mass spectrometry', *Molecular Systems Biology*, vol. 15, no. 9, e8994. <https://doi.org/10.15252/msb.20198994>

### Digital Object Identifier (DOI):

[10.15252/msb.20198994](https://doi.org/10.15252/msb.20198994)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Molecular Systems Biology

### General rights






Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# An integrated workflow for crosslinking mass spectrometry

Marta L Mendes<sup>1,†,‡</sup> , Lutz Fischer<sup>1,2,‡</sup> , Zhuo A Chen<sup>1</sup>, Marta Barbon<sup>3,4</sup> , Francis J O'Reilly<sup>1</sup> , Sven H Giese<sup>1</sup>, Michael Bohlke-Schneider<sup>1</sup>, Adam Belsom<sup>1,2</sup>, Therese Dau<sup>2</sup>, Colin W Combe<sup>2</sup>, Martin Graham<sup>2</sup>, Markus R Eisele<sup>5</sup>, Wolfgang Baumeister<sup>5</sup>, Christian Speck<sup>3,4</sup> & Juri Rappsilber<sup>1,2,\*</sup> 

## Abstract

We present a concise workflow to enhance the mass spectrometric detection of crosslinked peptides by introducing sequential digestion and the crosslink identification software xiSEARCH. Sequential digestion enhances peptide detection by selective shortening of long tryptic peptides. We demonstrate our simple 12-fraction protocol for crosslinked multi-protein complexes and cell lysates, quantitative analysis, and high-density crosslinking, without requiring specific crosslinker 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.

**Keywords** crosslinking mass spectrometry; protein–protein interactions; proteomics; software; structural biology

**Subject Categories** Methods & Resources; Proteomics

DOI 10.15252/msb.20198994 | Received 9 May 2019 | Revised 19 August 2019 | Accepted 21 August 2019

**Mol Syst Biol.** (2019) **15**: e8994

## Introduction

Crosslinking mass spectrometry (CLMS) has become a standard tool for the topological analysis of multi-protein complexes and has begun delivering high-density information on protein structures, insights into structural changes and the wiring of interaction networks *in situ* (O'Reilly & Rappsilber, 2018). The technological development currently focuses on enrichment strategies for crosslinked peptides and mass spectrometric data acquisition (Leitner *et al.*, 2013; Kolbowski *et al.*, 2017; Liu *et al.*, 2017), including newly designed crosslinkers (Kao *et al.*, 2011). MS2-cleavable crosslinkers, in particular, have celebrated recent

successes for the analysis of protein complexes (Wang *et al.*, 2017) or complex mixtures (Chavez *et al.*, 2013; Liu & Heck, 2015).

The focus on bespoke crosslinkers has left general steps of sample preparation, such as protein digestion, with less attention. Tryptic digestion generates crosslinked peptides of considerable size, a quality that has been exploited with their enrichment by SEC (Leitner *et al.*, 2012b), but one that poses as a potential problem regarding their detection. Replacing trypsin with proteases such as GluC, AspN and chymotrypsin does not change peptide size distributions fundamentally (Swaney *et al.*, 2010). 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 xiSEARCH, 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 therefore *a priori* druggable, however that hitherto have remained undisclosed by cryo-EM due to their flexible nature.

## Results

### Sequential digestion increases the number of identified crosslinks

We first tested this workflow on a standard mix of seven Bis[sulfosuccinimidyl] suberate (BS<sup>3</sup>) crosslinked 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

<sup>1</sup> Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany

<sup>2</sup> Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

<sup>3</sup> MRC London Institute of Medical Sciences (LMS), London, UK

<sup>4</sup> DNA Replication Group, Faculty of Medicine, Institute of Clinical Sciences (ICS), Imperial College London, London, UK

<sup>5</sup> Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, Martinsried, Germany

\*Corresponding author. Tel: +49 30 314-72374; E-mail: juri.rappsilber@tu-berlin.de

†These authors contributed equally to this work

‡Present address: Quantitative Biology Unit, Luxembourg Institute of Health, Luxembourg, Luxembourg

of 12 acquisitions, which is the protocol applied to all subsequent analyses presented here (Fig 1A). The results of this protocol for our standard proteins were compared to a parallel digestion using the same four enzymes and using trypsin alone in four replicates, maintaining the analytical effort comparable in all three cases (SEC fractionation, 12 injections). Sequential digestion produced the best results when compared to replica analyses and parallel digestion (Figs 1B and C, and EV1, Dataset EV1). Before assessing if this improvement translated into a gain of information in biological applications, we investigated the origin of the added data (Figs EV2 and EV3, Dataset EV4).

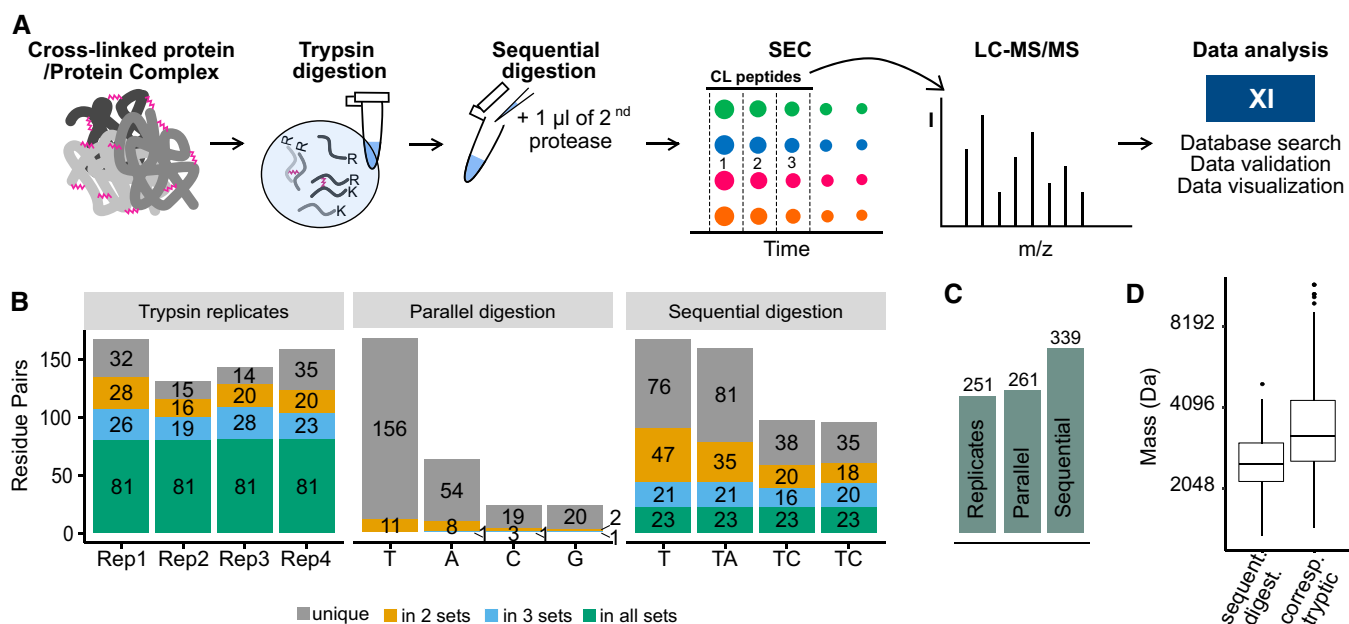
Indeed, sequential digestion led to smaller peptides than trypsin alone (Figs 1D and EV2F, and Dataset EV4) and moved the mass distribution of theoretical crosslinkable peptides more into the mass range typically detected by our instrument (Fig EV2F, Dataset EV4). For short peptides, we noticed a protection effect, based on the number of peptides containing missed cleavage sites and on the number of missed cleavage sites relative to peptide length (Figs EV3 and EV2B, and Dataset EV4). This agrees with reports that serine proteases lose efficiency as peptides shorten (Thompson & Blout, 1973; Wenzel & Tschesche, 1981). Although AspN is a metalloprotease, it showed a similar loss of efficiency for short peptides. Notably, we observed a bias towards maintaining tryptic C-termini. Crosslinked peptides with two tryptic C-termini are more frequently

identified while those with C-termini generated by the second protease are less frequent than expected, relying on N-termini as internal reference (Fig EV4). This identification bias is consistent with better fragmentation behaviour of peptides with basic C-termini (Olsen *et al*, 2004) and testifies to the importance of trypsin as part of the protocol.

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 (Chen *et al*, 2016c). Additionally, we tested the approach using two different crosslinkers, the homobifunctional crosslinker BS<sup>3</sup> and the heterobifunctional, photoactivatable crosslinker sulfo succinimidyl 4,4'-azipentanoate (SDA).

### Compatibility with photo- and quantitative CLMS

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/](http://predictioncenter.org/download_area/CASP12/extra_experiments/); Appendix Fig S1A) using SDA as crosslinker and 26 LC-MS runs (Ogorzalek *et al*,



**Figure 1. Sequential digestion workflow compared to repeated analysis and parallel digestion.**

- A Sequential digestion workflow. Proteins or protein complexes are crosslinked and digested with trypsin. After splitting the sample into four aliquots, one remains single digested with trypsin (T) while the others are sequentially digested with either AspN (A), chymotrypsin (C) or GluC (G). Samples are enriched by SEC, and the three high-MW fractions are analysed by LC-MS, submitted to xiSEARCH and xiFDR analysis.
- B Results of the sequential digestion workflow applied to a synthetic 7-protein mix, compared to using trypsin alone in four replicates and parallel digestion with trypsin, AspN, chymotrypsin and GluC. A trypsin four replicate experiment shows a large overlap of the four datasets with little gain. Parallel digestions with trypsin, AspN, chymotrypsin and GluC demonstrate high complementarity but moderate gains over trypsin. Sequential digestion shows low overlap between the four datasets and the largest gain in unique residue pairs.
- C Gains of repeated analysis (trypsin only), parallel digestion and sequential digestion for the same data as shown in panel (B).
- D Crosslinked peptides obtained by sequential digestion of a synthetic 7-protein mix, compared to using trypsin alone in four replicates and parallel digestion with trypsin, AspN, chymotrypsin and GluC. Crosslinked peptides obtained by sequential digestion are smaller than their corresponding tryptic peptides. Boxplot ranges represent the 25<sup>th</sup> (lower hinge) and 75<sup>th</sup> (upper hinge) percentiles, respectively. Middle line represents the median. For trypsin 4 replicates were analysed and for sequential digestion and parallel digestion 1 sample was analysed.

2018). Using sequential digestion, we now identified 1,523 unique residue pairs in only 12 runs (Appendix Fig S1B and C, Dataset EV1 and EV2). With 5% long-distance links ( $> 20 \text{ \AA}$ ) when mapped onto the structure released by CASP organisers (Appendix Fig S1D), the 300% 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; Schmidt *et al*, 2013; Tomko *et al*, 2015; Chen *et al*, 2016b) with our workflow (Appendix Fig S2A) to investigate the dimerisation of C3b. Thioester-mediated dimerisation 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 (Hong *et al*, 1991; Rawal & Pangburn, 2001; Pangburn & Rawal, 2002). 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 crosslinks, about three times more than with trypsin alone (99) (Appendix Fig S2B–D; Dataset EV1 and EV2) which lends robust support to a bottom-to-bottom orientation (Appendix Fig S2E). 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.

### A novel and functionally important contact in the OCCM complex

Turning our attention to protein complexes, we investigated the OCCM complex, a helicase loading intermediate formed during the initiation of DNA replication. Recently, a  $3.9\text{-\AA}$  structure of *Saccharomyces cerevisiae* OCCM on DNA was obtained by cryo-electron microscopy (cryo-EM), supported by CLMS (Yuan *et al*, 2017). We identified 682 residues pairs from the same sample analysed before, with large contribution from sequential digestion (Fig 2A and Appendix Fig S3A–C; Dataset EV1 and EV2). 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 2B, lane 5) and analyse its biological relevance in a well-established *in vitro* helicase loading assay, which recapitulates the *in vivo* process (Evrin *et al*, 2009). 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-hexamers (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 (Barbon *et al*, in preparation), a process that is only poorly understood. Moreover, the CLMS data show that the Mcm2 C-

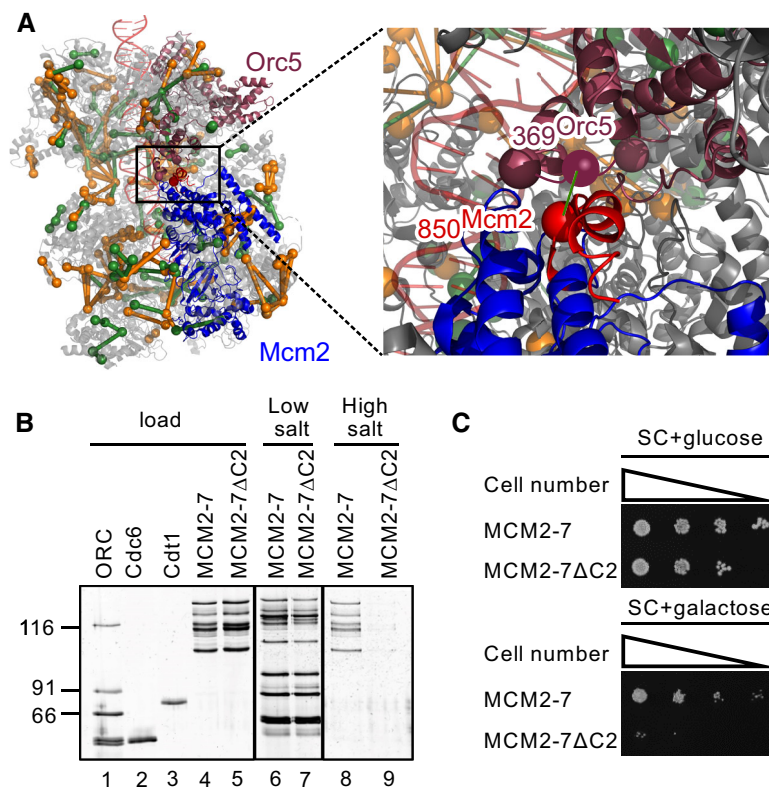
terminus is involved in a network of interactions with flexible domains of Orc6, Orc2 and Mcm5, indicating dynamics at the Mcm2-Mcm5 DNA entry gate (Samel *et al*, 2014), which could represent an ideal target for the development of inhibitors with potential as anti-cancer therapy (Gardner *et al*, 2017), as dynamic interactions have improved druggability characteristics over stable protein interactions (Ulucan *et al*, 2012; Jubb *et al*, 2015). Indeed, expressing Mcm2-7 $\Delta$ C2 causes dominant lethality (Fig 2C). The ability of CLMS data to complete the cryo-EM structure of the OCCM complex by providing information on dynamic contacts proved here essential. Note that 15% of our residue pairs falling into the published OCCM structure were long distance ( $> 30 \text{ \AA}$ , Appendix Fig S3D). 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.

### Conformational diversity of the 26S proteasome

We next analysed an affinity-purified 26S proteasome sample, containing more than 600 proteins (Dataset EV3). The results of our workflow compare favourably with the largest analysis reported on this complex to date (Wang *et al*, 2017) in terms of numbers ( $n = 1,644$  vs. 447 unique residue pairs in the proteasome at 5% FDR; Fig 3A and B, Appendix Fig S4A and B, Dataset EV1 and EV2) and in terms of agreeing with the structure of the individual subunits (6% vs. 26% long-distance links ( $> 30 \text{ \AA}$ ); Fig 3A and B). 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 lead to a large conformational variety (Fig 3C, Appendix Fig S4C and D). Indeed, some of these links ( $n = 78$ ) not matching to one structure of the proteasome mapped well to alternative conformational states stabilised by ATP analogue (Unverdorben *et al*, 2014; Wehmer *et al*, 2017). State-specific crosslinks were found predominant in the AAA-ATPase-dependent heterohexameric ring (Fig 3D–G Appendix Fig S4C and D) indicating rearrangement of Rpn5 relative to Rpt4 (Fig 3D). In the s2 state, our data support Rpn1 being translated and rotated to be positioned closer to the AAA-ATPase (Fig 3E). Crosslinks therefore support in solution the cryo-EM-based model of substrate transfer to the mouth of the AAA-ATPase heterohexameric ring (Unverdorben *et al*, 2014) 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.

### Exploration of the human cytoplasm

To probe our 12-fraction protocol in large-scale CLMS, we analysed seven high-molecular weight fractions of human cytosol. We identified 3,572 unique residue pairs (5% FDR, 528 proteins, Figs 4A, and EV5A and B, Dataset EV1 and EV2). This is in line with recent studies reporting 1,663 and 3,045 unique residue pairs, respectively,



**Figure 2. Sequential digestion of the affinity-purified complexes OCCM (*Saccharomyces cerevisiae*; 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.

albeit using a cleavable crosslinker and analysing whole cell extracts (Liu *et al.*, 2015, 2017). While the overlap between the published data and ours is low (Fig 4A), this will be influenced by factors such as the different starting material and the different analytical strategies.

Our protein–protein interaction network included previously observed complexes such as the Mcm2-7 complex, the 26S proteasome, the ribosome, the COPI complex, the TRiC-CCT complex and the HS90B-CDC37-Cdk4 complex (Figs 4B and EV5C–G, and Appendix Fig S5). For the 26S proteasome, we were able to distinguish between different states defining flexibility in the AAA-ATPase ring (Fig 4C; Chen *et al.*, 2016a). This indicates the ability of our protocol to unveil dynamic interactions in mixtures nearing the native environment complexity of proteins.

#### xiSEARCH, identification of crosslinks from mass spectra

To analyse the mass spectrometric data of these and other studies (Liu *et al.*, 2015, 2017; Wang *et al.*, 2017), we developed our database search software xiSEARCH (Figs 5A and EV6, Appendix Tables S1 and S2, <https://rappsilberlab.org/software/>). The algorithm of xiSEARCH has been described conceptually before (Giese *et al.*, 2016). It follows an approach that

computationally unlinks crosslinked peptides and by doing so circumvents the n2 database problem of crosslinking. Like pLink 2 (Chen *et al.*, 2019), StavroX (Götze *et al.*, 2012) and Kojak (Hoopmann *et al.*, 2015), xiSEARCH allows to search any cross-link and protease specificity; thus, xiSEARCH's performance was assessed against these three alternatives (Fig 5B). StavroX became non-responsive and was not pursued further. Kojak was paired with PeptideProphet (Keller *et al.*, 2002) and xiSEARCH with xiFDR (Fischer & Rappsilber, 2017) to maximise results and control the error rate. We assess the error on the level of unique residue pairs—as this is the actual information of interest. This is native to XiFDR while for Kojak(+PeptideProphet) and pLink 2 the output was sorted by score on the level of PSMs, only the best scoring PSM per residue pair was kept and a 5% FDR on the, now unique residue pairs calculated. By default, xiSEARCH weights the likelihood of a K vs. S, T or Y being involved in a crosslink higher to reduce the number of unique links without strong support by data. For the purpose of comparison, xiSEARCH was run with this feature enabled (marked xiSEARCH\*) and with this feature disabled, as none of the other tools are supporting a similar consideration. xiSEARCH reports 91% more unique links than Kojak + PeptideProphet and 45% more than pLink 2 (Fig 5B, Appendix Fig S6). Note that we and

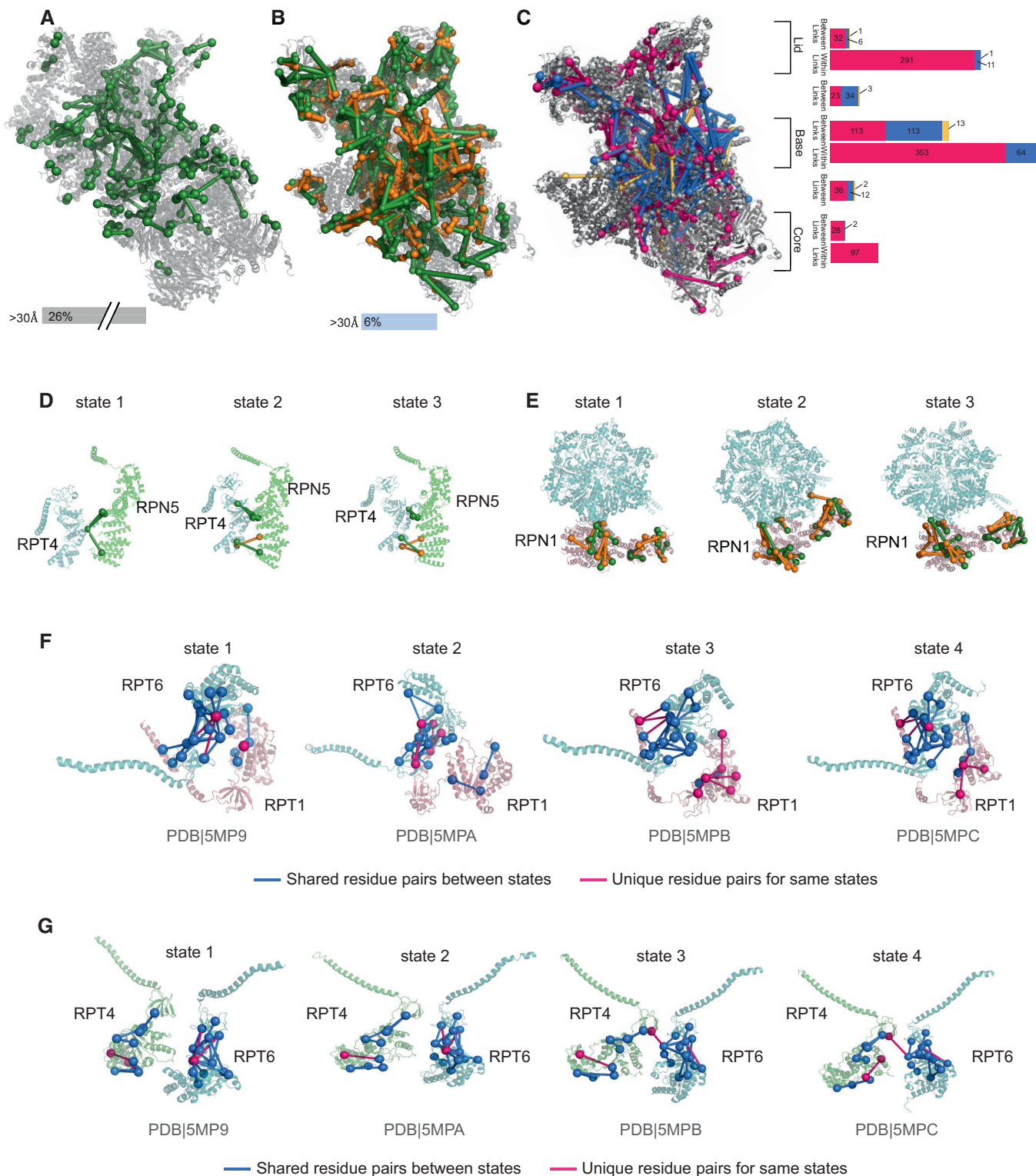


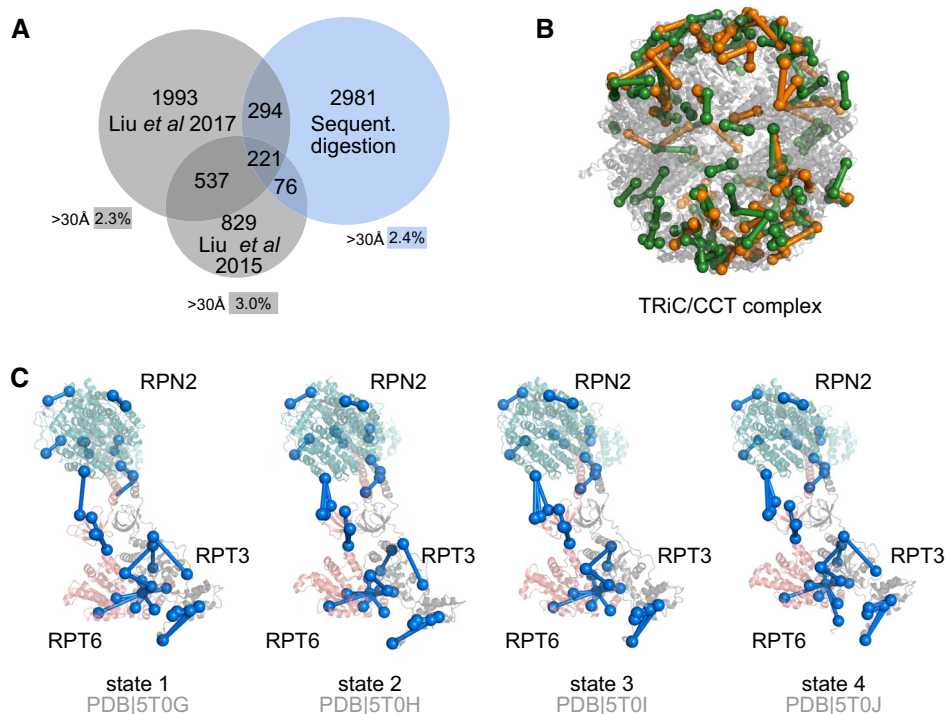
Figure 3.

the developers of these tools could not pair Kojak with Peptide-Prophet for sequential digestion. To check the reliability of residue pairs uniquely reported by xiSEARCH, we assessed them for the structurally rigid proteasome core particle. Less than 3%

were found to be long distance and thus very plausibly false, which is in good agreement with the expected FDR of 5%. In summary, xiSEARCH performed very favourably compared to other universal software for CLMS.

**Figure 3. Sequential digestion of the 26S proteasome from *Saccharomyces cerevisiae*.**

- A Unique residue pairs obtained by Wang *et al* for the human 26S proteasome (PDB 5GJR).
- B Unique residue pairs obtained by sequential digestion 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. Tryptic residue pairs are represented in green and non-tryptic in orange.
- C 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.
- D Unique residue pairs were mapped into the three states described by Unverdorben *et al* showing the rearrangement of Rpn5 relative to Rpt4.
- E Our data support Rpn1 being translated and rotated to be positioned closer to the AAA-ATPase.
- F Structural rearrangements of the AAA-ATPase-dependent heterohexameric ring throughout four states for the RPT6 and RPT1 mapped to the four states described by Wehmer *et al*.
- G Structural rearrangements of the AAA-ATPase-dependent heterohexameric ring throughout four states for the RPT4 and RPT6.

**Figure 4. Sequential digestion of the 26S proteasome from human cytosol.**

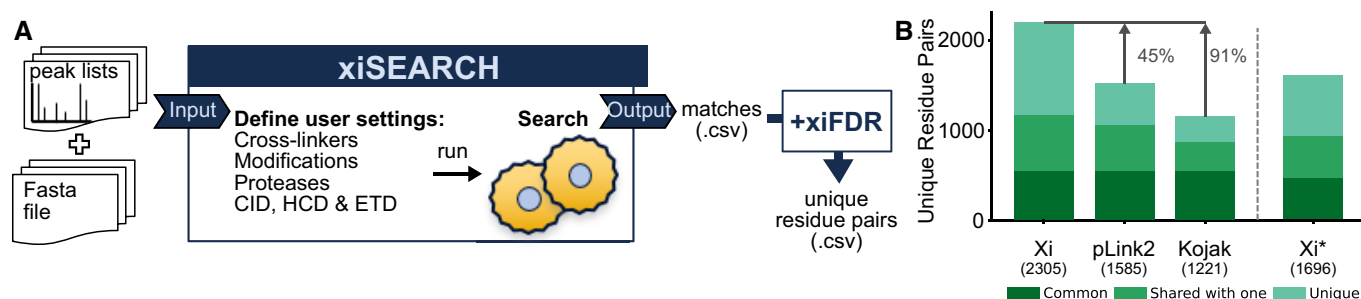
- A Comparison of sequential digestion for complex mixtures with previous studies. Sequential digestion returns a higher number of residue pairs with low overlap to published datasets, showing the complementarity of the different approaches. Long-distance links were determined only within proteins and are comparable for all datasets.
- B 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* (2012a) (PDB 4V94).
- C Despite the complexity of the sample, we were able to identify the four states of the 26S proteasome showing the flexibility of the AAA-ATPase-dependent heterohexameric ring.

## Discussion

### Sequential digestion novelty

The use of proteases other than trypsin to achieve complementary information in protein analysis dates back to at least 1987 (Aeberold *et al*, 1987). Since then, several works used parallel digestion and fewer used sequential digestion to increase sequence coverage and/or complementarity in simple protein complexes (Mohammed *et al*, 2008), proteomes (MacCoss *et al*, 2002; Swaney *et al*, 2010;

Guo *et al*, 2014), phosphoproteomes (Wang *et al*, 2008; Gilmore *et al*, 2012; Giansanti *et al*, 2015) and other post-translational modifications (Larsen *et al*, 2005). In CLMS, parallel digestion was reported first by Pinkse *et al* (2009) to increase complementarity and target a specific crosslink site. Leitner *et al* in 2012 presented crosslink data obtained when using in parallel five different proteases on a mix of standard proteins. Unfortunately, it remained unclear if the parallel use of five proteases would have been outperformed by simply five times re-analysing a tryptic digest. Re-analysing tryptic digests results in additional crosslinks being



**Figure 5. xiSEARCH.**

- A xiSEARCH is an open source search engine that takes a peak list as input. Users can define any type of crosslinker, modification, digestion and fragmentation method. The output is a list of matches in .csv format. We use xiFDR to filter results to the desired confidence level.
- B xiSEARCH + xiFDR(Xi), pLink 2 and Kojak(+PeptideProphet) comparison at 5% residue-pair FDR. The same trypsin dataset of the 26S proteasome was searched with all three software packages. xiSEARCH was run twice—once giving same likelihood for matching lysine, serine, threonine and tyrosine (xiSEARCH), as is the case for Kojak and pLink 2, and once giving priority to Lysine (Xi\*).

detected (Müller *et al*, 2018) also seen here (analysing a trypsin digest four times increased the observation from 167 to 251 unique residue pairs (URPs), Fig 1B and C). We then analytically prove that parallel digestion outperforms repeated injection of tryptic digests for crosslink detection, albeit moderately (261 vs. 251 URPs, Fig 1C). Importantly, we expand this by showing that sequential digestion is even better (339 URPs, Fig 1C).

We underpin these observations by proposing a mechanistic model that explains why sequential digestion outperforms parallel and repeated use of proteases. Outperforming repeated use of trypsin is linked to sequentially digested crosslinked peptides being smaller (Fig EV2A, Dataset EV4) which in fact improves detection rates as most observed peptides fall between 1,000 and 2,000 Da (Fig EV3D). The second digestion targets primarily long peptides as can be seen both in the surviving long peptides (poor in secondary cleavage sites) as well as in the observed short peptides (rich in secondary cleavage sites; Fig EV3). Consequently, sequential digestion is effective and does not shorten crosslinked peptides such that they become less observable (below 1,000 Da). Outperforming parallel digestion is linked to the detection bias of proteomics towards peptides with tryptic C-termini, that has been noted for linear peptides (Giansanti *et al*, 2016). Cleaving a crosslinked peptide N-terminal of the crosslink site maintains the tryptic C-termini while cleaving C-terminal leads to a non-tryptic C-terminus. During digestion, both should be equally likely. However, the former will be more likely to be detected which is reflected in our data (Fig EV1C). As all crosslinked peptides of parallel digestion other than those of trypsin use are lacking tryptic C-termini, parallel digestion has a marked disadvantage.

### Search software comparison

In our hands, xiSEARCH performs favourably when compared with other crosslink search software packages. However, a software comparison involving only the developers of one of the software packages will likely always be incomplete. A balanced evaluation of the merits of crosslink search tools would ideally be based on a community-wide effort with contributions of the developers of the respective software packages.

Here, we were restricted to software tools that supported sequential digestion and that supported a comparable set of features as used in our data analysis. One such point is that we search BS3 with specificity to lysine, serine, threonine and tyrosine. This already excludes many crosslink software packages, due to their restriction to lysine only. Furthermore, we place linkage sites in the absence of strong distinctive evidence preferentially on lysine. To our knowledge, this is not supported by any other crosslink search software. For the comparison, we removed that preference—likely inflating false results (Appendix Fig S6).

### Protocol applicability

Most published protocols are tested only for specific samples and application areas. Furthermore, workflows tend to build on one specific combination of crosslinker and database search tool. In contrast, we deliver a protocol with broad applicability, demonstrating its use in protein dynamics, protein complexes topology, conformational changes (qCLMS) and using homo- and hetero-bifunctional crosslinkers. Our integrated workflow utilises standard crosslinkers, without special chemistries to assist analysis. This permits the sequential digestion workflow to be combined with other crosslinkers such as MS-cleavable crosslinkers. Recent large-scale studies have successfully used MS-cleavable crosslinkers (Liu *et al*, 2015; Hage *et al*, 2017). Note, however, that our work uses standard crosslinkers at no obvious disadvantage. Importantly, MS-cleavable crosslinkers have yet to be combined with high-density crosslinking and are likely incompatible with crosslinking by non-canonical amino acids (Suchanek *et al*, 2005) motivating efforts in keeping crosslink 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. In particular, we were able to identify dynamic protein interaction regions and topologies which are notoriously difficult to detect using conventional structural biology methods due to their flexibility yet are prime therapeutic intervention points, as these important interactions are only short-lived and therefore druggable.



### Complementarity of CLMS to cryo-electron microscopy

Crystallography and cryo-EM employ averaging approaches to obtain high-resolution structural information of proteins or protein complexes. Structural heterogeneity is not compatible with crystallography, while novel cryo-EM data analysis approaches can deal with some conformational heterogeneity. Nevertheless, cryo-EM also fails to visualise flexible regions.

Indeed, in the context of the OCCM, cryo-EM revealed the overall organisation of the 14 subunit complex. However, in an area of high protein flexibility, which encompasses the important Mcm2/Mcm5 DNA entry gate of the MCM2-7 ring, the EM structure was hampered by low resolution. By CLMS, a new network of interactions was detected at the DNA entry gate, involving Mcm2, Mcm5, Orc2 and Orc6. Moreover, we showed that the Mcm2 interaction with Orc6 is functionally relevant, as a small Mcm2 deletion in the

interaction surface strongly affected helicase loading and caused lethality *in vivo*. As such, CLMS and cryo-EM can synergise to identify and characterise flexible regions in protein complexes, which have important functionality.

Interestingly, protein–protein interactions identified by cryo-EM and crystallography are nearly impossible to target by small molecules, since they are very long-lived and due to their hydrophobicity not accessible for water-soluble small molecules. On the other hand, CLMS, employing water-soluble crosslinkers, naturally identifies well-hydrated dynamic protein–protein interaction surfaces, which are typically less stable or context-dependent and therefore ideally suited for drug development. In summary, CLMS is uniquely capable of detecting flexible and dynamic protein interactions, which makes the technology highly synergistic with other structural approaches and opens a window of opportunity for drug development of dynamic protein–protein interactions.

## Materials and Methods

### Reagents and Tools table

Reagent/resource	Reference or source	Identifier or catalogue number
<b>Experimental models</b>		
YYS40 ( <i>S. cerevisiae</i> )	Sakata, E. <i>et al</i>	Prof. Dr. Wolfgang Baumeister, MPI, Martinsried
K562 cells ( <i>Homo sapiens</i> )	DSMZ	#ACC-10
OCCM complex	Evrin, C. <i>et al</i>	Prof. Dr. Christian Speck, MRC-LMS or ICS-ICL, London
<b>Proteins</b>		
Catalase	SIGMA	C9322
Myoglobin	SIGMA	M1882
Cytochrome C	SIGMA	C2037
Lysozyme	SIGMA	L6876
Human serum albumin	SIGMA	A8763
Conalbumin	SIGMA	C0755
Creatine kinase	Roche	10127566001
C3b	Complement Technology, Inc.	A114
UGGT	CASP 12	Prof. Pietro Roversi
<b>Chemical, enzymes and other reagents</b>		
Bis[sulfo succinimidyl]suberate	Thermo Fisher	26173
sulfosuccinimidyl 4,4'-azipentanoate	Thermo Fisher	21580
Trypsin	Thermo Fisher	90057
AspN	Promega	V1621
Chymotrypsin	Promega	V1061
GluC	Promega	V1651
NuPAGE™ Novex 4–12% Bis-Tris	Thermo Fisher	NP0335BOX
Dithiothreitol	Merck	1114740005
Iodoacetamide	SIGMA	I1149
Ammonium bicarbonate	SIGMA	A6141
TFA	SIGMA	T6508
Formic acid	Fluka	94318
Acetonitrile	Riedel-de Haen	34967

Reagents and Tools table (continued)

Reagent/resource	Reference or source	Identifier or catalogue number
<b>Equipment</b>		
Superdex Peptide 3.2/300	GE Healthcare	GE29-0362-31
Shimadzu HPLC	Shimadzu	
EASY-Spray™ LC column	Thermo Fisher	ES803
Dionex Ultimate 3000 RSLCnano	Thermo Fisher	
Orbitrap Fusion Lumos Tribrid	Thermo Fisher	

## Methods and Protocols

### Sample preparation, crosslinking and digestion with trypsin

The seven standard proteins catalase, myoglobin, cytochrome C, lysozyme, creatine kinase, HSA and conalbumin were resuspended in BS<sup>3</sup> crosslinking buffer (20 mM HEPES, 20 mM NaCl, 5 mM MgCl<sub>2</sub>, pH 7.8) to a final concentration of 1 mg/ml. Crosslinker was added to a 1:1 (*w/w*) protein to crosslinker ratio and samples incubated for 2 h on ice. Crosslinking reaction was quenched with excess ammonium bicarbonate (ABC) for 1 h at room temperature (RT). The seven crosslinked proteins were loaded on NuPAGE™ 4–12% Bis-Tris protein gels to isolate the monomeric band of each protein that was then extracted, and in-gel digested with trypsin (Shevchenko *et al*, 2006). After peptide extraction from the gel, peptides from each protein were mixed in a 1:1 weight ratio to a final amount of 200 µg divided into four parts and desalted using C18-StageTips (Rappsilber *et al*, 2003).

For the OCCM complex, pUC19-ARS1 beads were used to assemble the OCCM complex as described elsewhere (Evrin *et al*, 2009). Crosslinking was performed on beads. BS<sup>3</sup> was added to 200 µg of the OCCM complex to a 1:8,100 protein to crosslinker molar ratio. The sample was incubated for 2 h on ice, and the crosslinking reaction was quenched with excess ABC for 1 h 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. Reaction was stopped with 10% (*v/v*) trifluoroacetic acid (TFA), and the samples were divided into four parts and desalted using StageTips.

The 26S proteasome was isolated from *S. cerevisiae* by affinity purification using the 3× FLAG-tagged subunit Rpn11 as described elsewhere (Sakata *et al*, 2011). For the crosslinking, the 26S proteasome buffer was exchanged to BS<sup>3</sup> crosslinking buffer using 30 kDa molecular weight cut-off (MWCO) filters (Millipore). 200 µg of the 26S proteasome was crosslinked with BS<sup>3</sup>. BS<sup>3</sup> was added to a 1:1 (*w/w*) protein to crosslinker ratio. Samples were incubated for 2 h on ice, and the crosslinking 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 were divided into four parts and desalted using C18-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<sub>2</sub> incubators in RPMI 1640 media supplemented with 10% (*v/v*) foetal bovine serum (FBS) + 2 mM glutamine. 3 × 10<sup>8</sup> cells were harvested by centrifugation (180 × *g*) and washed three 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<sub>2</sub>, 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 × *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 × 600) size exclusion column on an Äkta Purifier (HPLC) system running at 0.25 ml/min 100 mM HEPES pH 7.2, 100 mM KCl, 20 mM NaCl, 3 mM MgCl<sub>2</sub>, 1 mM EDTA and 10% glycerol. 500 µl fractions were collected. Protein fractions were crosslinked at < 1 mg/ml (quantified by Bradford) and with 2:1 *w/w* ratio of BS<sup>3</sup> to protein. Samples were *in-solution* digested with trypsin as described above. Samples were divided into four parts and desalted using C18-StageTips.

C3b monomer and dimer were labelled as described elsewhere (Chen *et al*, 2016c) and in-gel digested with trypsin. Samples were divided into four parts and desalted using C18-StageTips.

UGGT was crosslinked using sulfo-SDA using eight different proteins to crosslinker 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*)]. Crosslinking was carried out in two stages: firstly, sulfo-SDA, dissolved in SDA crosslinking buffer (25 µl, 20 mM HEPES-OH, 20 mM NaCl, 5 mM MgCl<sub>2</sub>, pH 7.8), was added to target protein (25 µg, 1 µg/µl) and left to react in the dark for 50 min 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 min. 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 protein was *in-solution* digested with trypsin as described above. Sample was divided into four parts and desalted using C18-StageTips.

For all models, one part of the sample digested with trypsin was fractionated by SEC and the remaining three parts were sequentially digested with AspN, chymotrypsin and GluC, respectively.

### Parallel digestion

Parallel digestion with AspN, chymotrypsin and GluC of the seven standard proteins was performed as described for trypsin. After isolating the monomeric bands of each one of the seven standard proteins, those were in-gel digested in parallel with AspN, chymotrypsin and GluC. After peptide extraction from the gel, peptides from each protein digested with the same protease were mixed in a 1:1 weight ratio to a final amount of 200 µg resulting in four samples of the protein standards, one digested with AspN, one digested with chymotrypsin and one digested with GluC. 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 parallel digestion, samples were fractionated by SEC and analysed by LC-MS/MS.

### Sequential digestion

After desalting of tryptic peptides, sequential digestion was performed as follows:

- 1 Resuspend tryptic peptides in ABC 50 mM.
- 2 For sequential digestion with AspN:
  - 2.1 Add AspN to a final protease:protein ratio of 1:100 (*w/w*).
  - 2.2 Incubate ON at 37°C.
  - 2.3 Acidify samples using 10% TFA.
  - 2.4 Reduce sample volume to 50 µl by evaporation using a vacuum concentrator.
  - 2.5 Fractionate by SEC.
- 3 For sequential digestion with chymotrypsin:
  - 3.1 Add chymotrypsin to a final protease:protein ratio of 1:50 (*w/w*).
  - 3.2 Incubate ON at RT.
  - 3.3 Acidify samples using 10% TFA.
  - 3.4 Reduce sample volume to 50 µl by evaporation using a vacuum concentrator.
  - 3.5 Fractionate by SEC.
- 4 For sequential digestion with GluC:
  - 4.1 Add GluC to a final protease:protein ratio of 1:50 (*w/w*).
  - 4.2 Incubate ON at 37°C.
  - 4.3 Acidify samples using 10% TFA.
  - 4.4 Reduce sample volume to 50 µl by evaporation using a vacuum concentrator.
  - 4.5 Fractionate by SEC.

### Fractionation of peptides by size exclusion chromatography

Size exclusion chromatography (SEC) of crosslinked peptides was performed as described elsewhere (Leitner *et al*, 2013). 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 2 min over one column volume. The three high-MW fractions were dried, resuspended in 0.1% (*v/v*) TFA and analysed by LC-MS/MS. All samples in this work, including all the samples used in our proof of principle experiments—replicates, parallel and sequential digestions—of the seven standard proteins, were fractionated by SEC as described in our workflow in Fig 1A.

### In vitro pre-RC assay

The *in vitro* pre-RC assay was performed as described (Evrin *et al*, 2009; Fernández-Cid *et al*, 2013). 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 DNA 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-X100, 5% glycerol). After 20 min 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<sub>2</sub>.

### 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 h. Cells were grown in suspension to 10<sup>7</sup> cells/ml. Three microlitres of a fivefold serial dilution was 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 of 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 1,700 *m/z* and AGC target of 2.5 × 10<sup>5</sup> 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 5.0 × 10<sup>4</sup> were selected for fragmentation. Selected precursors were fragmented by HCD using a collision energy setting of 30%. MS2 spectra were acquired at a resolution of 15,000 and AGC of 10<sup>4</sup> with a maximum injection time of 35 ms. Dynamic exclusion was set to 60 s after one 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 the 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 xiSEARCH for identification of crosslinked peptides (xiSEARCH software is available from <https://rappsilbe>

rlab.org/downloads/and the code is available from <https://github.com/Rappsilber-Laboratory/xiSEARCH>). For the crosslinking 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 BS<sup>3</sup> crosslinked samples, carbamidomethylation of cysteines was set as a fixed modification and oxidation of methionine, the hydrolysed crosslinker (BS-OH: 156.0786 Da) and the amidated crosslinker (BS3-NH2: 155.0964 Da) were set as variable modifications. Reaction specificity for BS<sup>3</sup> (mass modification: 138.0681 Da) was assumed to be with Lysine, Serine, Threonine and Tyrosine or the protein N-terminus. For SDA crosslinked samples, carbamidomethylation of cysteines, oxidation of methionine and the crosslinker alone (mass modification: 109.0396 Da), hydrolysed (SDA-OH: 100.0524 Da) or crosslinker loop (SDA-loop: SDA crosslink within a peptide that is also crosslinked 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, 1a06 and 1ova); for C3b, we used the FASTA corresponding to the UniProt accession P01024; for the UGGT, the UniProt sequence of the protein was used (G0SB58); for the OCCM complex, a database containing all the 14 OCCM subunits was used; and for the 26S proteasome, a linear search of the data was first performed using a complete *S. cerevisiae* database (UniProt, release-2016\_11) and the proteins that were present at < 1% of the most abundant protein (as judged by iBAQ values) were excluded. Our rationale for excluding the least abundant proteins from crosslink searches was that crosslinks tend to be less easily identified than linear peptides and thus will be detected only in the more abundant proteins. Although likely not contributing detectable crosslinks, less abundant proteins will nevertheless contribute random matches and thus reduce the sensitivity of the search. For the human cytosol, linear searches were performed for each one of the initial SEC fractions using the entire human database (UniProt, release-2016\_02). Proteins that were present at < 1% (SEC fraction 1) or 0.5% (SEC fractions 2–7) of the most abundant protein (as judged by iBAQ values) were excluded. The different threshold was used to result in databases of about equal size. FDR was estimated on a 5% residue level, including only unique PSMs and boosting results, using xiFDR (Fischer & Rappsilber, 2017).

Cleavage site protection was calculated by dividing the mean of available miss-cleavages for the second protease in the sequential digestion datasets by the mean of the potential miss-cleavages in the trypsin dataset.

For software comparison, the 26S proteasome was searched using xiSEARCH (version 1.6.731), Kojak (version 1.5.5) and pLink 2 (version 2.3.2) using the same parameters as described before with minor alterations: Searches were performed using two missed cleavages, MS accuracy 3 ppm and MS/MS accuracy 20 ppm. For xiSEARCH, we used missing monoisotopic peaks 3. For xiSEARCH, Kojak and pLink 2, we gave the same preference for linkage sites in K, S, T and Y.

## Data availability

The datasets and computer code produced in this study are available in the following databases:

- Raw-files and identifications: ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (Vizcaíno et al, 2014) at: <https://www.ebi.ac.uk/pride/archive/projects/PXD008550>.
- The source code of xiSEARCH is available at: <https://github.com/Rappsilber-Laboratory/XiSearch>
- The source code of xiFDR is available at: <https://github.com/Rappsilber-Laboratory/xiFDR>

**Expanded View** for this article is available online.

## Acknowledgements

We thank Swantje Lenz for the technical support and helpful discussions and Petra Ryl for sample preparation support. We thank Pietro Roversi and CASP12 organisers for supplying UGGT sample and Alberto Riera for contributing OCCM sample. We thank Michael Hoopmann for his support with pairing kojak and peptideProphet. 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

MLM, LF and JR developed the study; MLM, ZAC, FJO'R, AB, MB and MRE performed sample preparation; MLM performed LC-MS analysis; MLM, ZAC, FJO'R, AB, TD, MB, CS, SG and MB-S performed data analysis; LF developed xiSEARCH; CWC and MG provided critical support in data analysis; MB, CS, MRE and WB contributed with data interpretation; and MLM, FJO'R, MB-S, LF and JR wrote the article with critical input from all authors.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SB (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after *in situ* protease digestion on nitrocellulose. *Proc Natl Acad Sci USA* 84: 6970–6974
- Chavez JD, Weisbrod CR, Zheng C, Eng JK, Bruce JE (2013) Protein interactions, post-translational modifications and topologies in human cells. *Mol Cell Proteomics* 12: 1451–1467
- Chen S, Wu J, Lu Y, Ma Y-B, Lee B-H, Yu Z, Ouyang Q, Finley DJ, Kirschner MW, Mao Y (2016a) Structural basis for dynamic regulation of the human 26S proteasome. *Proc Natl Acad Sci USA* 113: 12991–12996
- Chen ZA, Pellarin R, Fischer L, Sali A, Nilges M, Barlow PN, Rappsilber J (2016b) Structure of complement C3(H2O) revealed by quantitative cross-linking/mass spectrometry and modeling. *Mol Cell Proteomics* 15: 2730–2743

- Chen Z, Fischer L, Tahir S, Bukowski-Wills J-C, Barlow P, Rappsilber J (2016) Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes. *Wellcome Open Res* 1: 5
- Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, Liu C, Zeng W-F, Ding Y-H, Tan D et al (2019) A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* 10: 3404
- Evvin C, Clarke P, Zech J, Lurz R, Sun J, Uhle S, Li H, Stillman B, Speck C (2009) 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
- Fernández-Cid A, Riera A, Tognetti S, Herrera MC, Samel S, Evvin C, Winkler C, Gardenal E, Uhle S, Speck C (2013) 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
- Fischer L, Rappsilber J (2017) Quirks of error estimation in cross-linking/mass spectrometry. *Anal Chem* 89: 3829–3833
- Gardner NJ, Gillespie PJ, Carrington JT, Shanks EJ, McElroy SP, Haagensen EJ, Frearson JA, Woodland A, Blow JJ (2017) 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
- Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJR (2015) An augmented multiple-protease-based human phosphopeptide atlas. *Cell Rep* 11: 1834–1843
- Giansanti P, Tsiatsiani L, Low TY, Heck AJR (2016) Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* 11: 993
- Giese SH, Fischer L, Rappsilber J (2016) A study into the collision-induced dissociation (CID) behavior of cross-linked peptides. *Mol Cell Proteomics* 15: 1094–1104
- Gilmore JM, Kettenbach AN, Gerber SA (2012) Increasing phosphoproteomic coverage through sequential digestion by complementary proteases. *Anal Bioanal Chem* 402: 711–720
- Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kühn U, Sinz A (2012) StavroX—a software for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom* 23: 76–87
- Guo X, Trudgian DC, Lemoff A, Yadavalli S, Mirzaei H (2014) Confetti: a multiprotease map of the HeLa proteome for comprehensive proteomics. *Mol Cell Proteomics* 13: 1573–1584
- Hage C, Falvo F, Schäfer M, Sinz A (2017) Novel concepts of MS-cleavable cross-linkers for improved peptide structure analysis. *J Am Soc Mass Spectrom* 28: 2022–2038
- Hong K, Kinoshita T, Pramoonjago P, Kim YU, Seya T, Inoue K (1991) Reconstitution of C5 convertase of the alternative complement pathway with isolated C3b dimer and factors B and D. *J Immunol* 146: 1868–1873
- Hoopmann MR, Zelter A, Johnson RS, Riffle M, MacCoss MJ, Davis TN, Moritz RL (2015) Kojak: efficient analysis of chemically cross-linked protein complexes. *J Proteome Res* 14: 2190–2198
- Jubb H, Blundell TL, Ascher DB (2015) Flexibility and small pockets at protein-protein interfaces: new insights into druggability. *Prog Biophys Mol Biol* 119: 2–9
- Kao A, Chiu C-L, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L (2011) 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
- Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 74: 5383–5392
- Kolbowski L, Mendes ML, Rappsilber J (2017) Optimizing the parameters governing the fragmentation of cross-linked peptides in a Tribrid mass spectrometer. *Anal Chem* 89: 5311–5318
- Larsen MR, Højrup P, Roepstorff P (2005) Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Mol Cell Proteomics* 4: 107–119
- Leitner A, Joachimiak LA, Bracher A, Mönkemeyer L, Walzthoeni T, Chen B, Pechmann S, Holmes S, Cong Y, Ma B et al (2012a) The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* 20: 814–825
- Leitner A, Reischl R, Walzthoeni T, Herzog F, Bohn S, Förster F, Aebersold R (2012b) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics* 11: M111.014126
- Leitner A, Walzthoeni T, Aebersold R (2013) 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
- Liu F, Heck AJR (2015) Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr Opin Struct Biol* 35: 100–108
- Liu F, Rijkers DTS, Post H, Heck AJR (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* 12: 1179–1184
- Liu F, Lössl P, Scheltema R, Viner R, Heck AJR (2017) Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat Commun* 8: 15473
- MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A et al (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA* 99: 7900–7905
- Mohammed S, Lorenzen K, Kerkhoven R, van Breukelen B, Vannini A, Cramer P, Heck AJR (2008) Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem* 80: 3584–3592
- Müller F, Fischer L, Chen ZA, Auchynnikava T, Rappsilber J (2018) On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J Am Soc Mass Spectrom* 29: 405–412
- Ogorzalek TL, Hura GL, Belsom A, Burnett KH, Kryshatovych A, Tainer JA, Rappsilber J, Tsutakawa SE, Fidelis K (2018) Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. *Proteins* 86(Suppl 1): 202–214
- Olsen JV, Ong S-E, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 3: 608–614
- O'Reilly FJ, Rappsilber J (2018) Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* 25: 1000–1008
- Pangburn MK, Rawal N (2002) Structure and function of complement C5 convertase enzymes. *Biochem Soc Trans* 30: 1006–1010
- Pinkse MWH, Rijkers DTS, Dostmann WR, Heck AJR (2009) Mode of action of cGMP-dependent protein kinase-specific inhibitors probed by photoaffinity cross-linking mass spectrometry. *J Biol Chem* 284: 16354–16368
- Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75: 663–670
- Rawal N, Pangburn MK (2001) Structure/function of C5 convertases of complement. *Int Immunopharmacol* 1: 415–422

- Sakata E, Stengel F, Fukunaga K, Zhou M, Saeki Y, Förster F, Baumeister W, Tanaka K, Robinson CV (2011) The catalytic activity of Ubp6 enhances maturation of the proteasomal regulatory particle. *Mol Cell* 42: 637–649
- Samel SA, Fernández-Cid A, Sun J, Riera A, Tognetti S, Herrera MC, Li H, Speck C (2014) A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2–7 onto DNA. *Genes Dev* 28: 1653–1666
- Schmidt C, Zhou M, Marriott H, Morgner N, Politis A, Robinson CV (2013) Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nat Commun* 4: 1985
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1: 2856–2860
- Suchanek M, Radzikowska A, Thiele C (2005) Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* 2: 261–267
- Swaney DL, Wenger CD, Coon JJ (2010) Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 9: 1323–1329
- Thompson RC, Blout ER (1973) Dependence of the kinetic parameters for elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Biochemistry* 12: 57–65
- Tomko Jr RJ, Taylor DW, Chen ZA, Wang H-W, Rappsilber J, Hochstrasser M (2015) A single  $\alpha$  helix drives extensive remodeling of the proteasome lid and completion of regulatory particle assembly. *Cell* 163: 432–444
- Ulucan O, Eyrich S, Helms V (2012) Druggability of dynamic protein-protein interfaces. *Curr Pharm Des* 18: 4599–4606
- Unverdorben P, Beck F, Śledź P, Schweitzer A, Pfeifer G, Plitzko JM, Baumeister W, Förster F (2014) Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc Natl Acad Sci USA* 111: 5544–5549
- Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N et al (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 32: 223–226
- Wang B, Malik R, Nigg EA, Körner R (2008) Evaluation of the low-specificity protease elastase for large-scale phosphoproteome analysis. *Anal Chem* 80: 9526–9533
- Wang X, Cimermancic P, Yu C, Schweitzer A, Chopra N, Engel JL, Greenberg CH, Huszagh AS, Beck F, Sakata E et al (2017) Molecular details underlying dynamic structures and regulation of the human 26S proteasome. *Mol Cell Proteomics* 16: 840–854
- Wehmer M, Rudack T, Beck F, Aufderheide A, Pfeifer G, Plitzko JM, Förster F, Schulten K, Baumeister W, Sakata E (2017) Structural insights into the functional cycle of the ATPase module of the 26S proteasome. *Proc Natl Acad Sci USA* 114: 1305–1310
- Wenzel HR, Tschesche H (1981) Cleavage of peptide-4-nitroanilide substrates with varying chain length by human leukocyte elastase. *Hoppe Seylers Z Physiol Chem* 362: 829–831
- Yuan Z, Riera A, Bai L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B et al (2017) Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol* 24: 316–324



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.