Do Not Waste Time Ensure Success in Your Cross-Linking Mass Spectrometry Experiments before You Begin

Lucienne Nouchikian, David Fernandez-Martinez, Pierre-Yves Renard, Cyrille Sabot, Guillaume Duménil, Martial Rey, Julia Chamot-Rooke

analytical chemistry

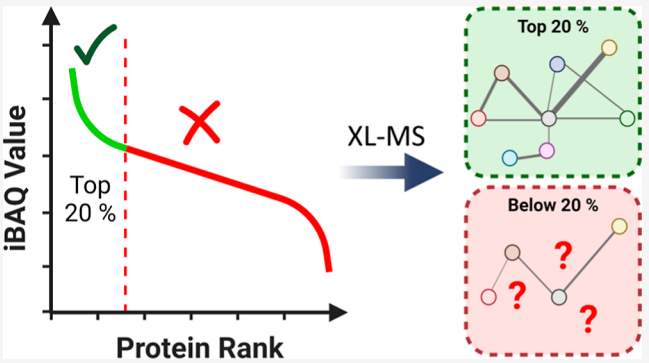
pubs.acs.org/ac
Article

Do Not Waste Time—Ensure Success in Your Cross-Linking Mass Spectrometry Experiments before You Begin

Lucienne Nouchikian, David Fernandez-Martinez, Pierre-Yves Renard, Cyrille Sabot, Guillaume Duménil, Martial Rey, and Julia Chamot-Rooke*

Cite This: *Anal. Chem.* 2024, 96, 2506−2513
Read Online

ABSTRACT: Cross-linking mass spectrometry (XL-MS) has become a very useful tool for studying protein complexes and interactions in living systems. It enables the investigation of many large and dynamic assemblies in their native state, providing an unbiased view of their protein interactions and restraints for integrative modeling. More researchers are turning toward trying XL-MS to probe their complexes of interest, especially in their native environments. However, due to the presence of other potentially higher abundant proteins, sufficient cross-links on a system of interest may not be reached to achieve satisfactory structural and interaction information. There are currently no rules for predicting whether XL-MS experiments are likely to work or not; in other words, if a protein complex of interest will lead to useful XL-MS data. Here, we show that a simple iBAQ (intensity-based absolute quantification) analysis performed from trypsin digest data can provide a good understanding of whether proteins of interest are abundant enough to achieve successful cross-linking data. Comparing our findings to large-scale data on diverse systems from several other groups, we show that proteins of interest should be at least in the top 20% abundance range to expect more than one cross-link found per protein. We foresee that this guideline is a good starting point for researchers who would like to use XL-MS to study their protein of interest and help ensure a successful cross-linking experiment from the beginning. Data are available via ProteomeXchange with identifier PXD045792.



Cross-linking mass spectrometry (XL-MS) has become a well-integrated technique in the world of structural biology and systems biology.[1−3] In XL-MS, a chemical linker covalently binds with amino acids that are in close proximity. Subsequent digestion of the linked proteins and MS analysis allow us to identify the linked proteins and attain amino acid level resolution as to where the linkage occurred. As cross-linkers have known spacer lengths, the distance constraints they provide has made XL-MS a valuable resource for structural characterization of proteins and their assemblies, including highly dynamic systems with many interactions.[4−6] XL-MS data is frequently used in conjunction with data from other structural techniques, notably X-ray crystallography and cryo-electron microscopy to provide complementary information,[7,8] such as additional protein states,[9] filling in structural resolution gaps,[10] and assist in protein identification of unknown densities.[11] Cross-linking data is also used in computational modeling[2,12−14] to either improve on existing proteins models or through *de novo* modeling if a structure from the PDB or a suitable model is not available, with new tools emerging to aid in more realistic protein modeling, taking into consideration protein energy landscapes.[15−17] Another major application of XL-MS is identifying protein−protein interactions (PPIs),[18−20] especially as the field has quickly moved toward complex systems and in vivo (or in-cell) cross-linking. In vivo cross-linking allows for a global overview of protein interactions without prior assumptions[21] as it preserves potential transient interactions and offers the possibility of detecting novel PPIs that may be lost during purification.

Although many in vivo studies have been conducted with a variety of different cell types and cross-linkers, majority of the structural and interaction information arises from highly abundant proteins, especially from the cytosolic component, such as RNA polymerase, ribosomes, and proteasomes, but also on more highly abundant membrane protein assemblies, such as ATP synthase.[4] The fact that cross-linking favors proteins of higher abundance is established,[22] making information on lower abundant assemblies more limited in complex samples. Many in vivo studies have provided good cross-link proteome coverage from high to low abundance

Received: October 17, 2023
Revised: January 9, 2024
Accepted: January 19, 2024
Published: January 31, 2024

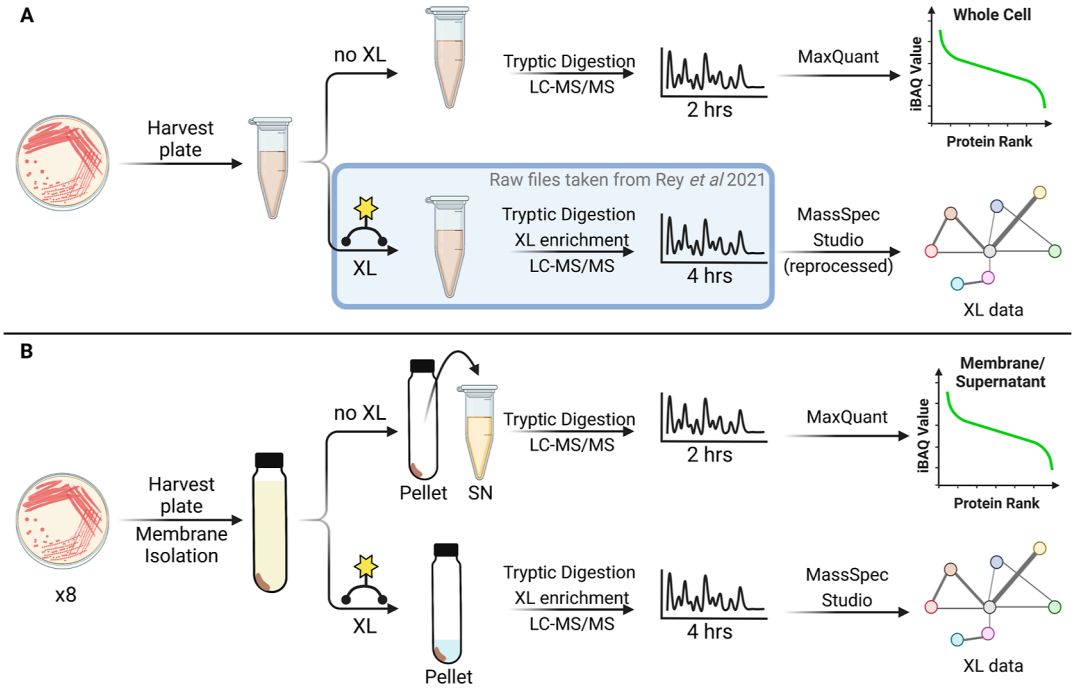


Figure 1. Summary of experiments performed on (A) the whole cell and (B) the membrane pellet isolation of *N. meningitidis*.

proteins, achieving cross-links from all cellular compartments.[21,23−26] However, if the goal of the experiment is to study the structure and interaction partners of a specific protein assembly, then the interference of other higher abundant proteins may result in insufficient cross-linking information for that system. A good balance must be reached where enough cross-linking information is captured for structural analysis and creation of informative PPI networks while still preserving a native environment and as many protein interactors as possible.

Studies have tried various ways to incorporate protein abundance information into cross-link identification. One study restricted the database to the top 40 proteins with the highest spectral counts from affinity pull-down enrichments.[27] Other studies have used iBAQ (intensity-based absolute quantification) values, which are often used in bottom-up proteomic experiments to determine the dynamic range of protein abundances in a sample. These values are then typically used to establish a "cut-off" point to simplify the protein database. It is reasoned that it is more difficult to identify cross-links than linear peptides in less abundant proteins[22] and helps remove false positives while speeding up the search. Different methods, such as removing the last 5%,[18] 1%,[22] after the first inflection point,[28] or taking the top 400 proteins,[29] have been used, showing that there is no clear consensus as to what is considered "abundant enough" to achieve acceptable XL-MS data. What is required here is a simple and general method to gauge (i) if the system of interest is sufficiently abundant or (ii) if an required enrichment step is sufficient for an XL-MS experiment. This is what we propose in the present paper.

Previously, our group introduced a novel trifunctional cross-linker, NNP9,[30] that can be used to bio-orthogonally enrich cross-linked peptides using click chemistry. We demonstrated its effective use in vivo on *Neisseria meningitidis*.[23] Yet, as with other studies, we noticed that most of the cross-link information was primarily dominated by abundant cytosolic proteins such as ribosomes or proteins involved in metabolic pathways. Information on proteins and their complexes which could bring more relevant and interesting biological results, such as on *Neisseria*'s many virulence factors, were considerably sparse. Interactions from only one such complex between three highly abundant proteins (PorB, RmpM, and FbpA) responsible for iron uptake were described, yet XL-MS data on its many other membrane-associated virulence factors[31] was overshadowed.

Starting from this data, we wondered how deep we could go in the proteome to obtain sufficient cross-link information on specific complexes and at which point would proteins be low enough in abundance where we could expect little to no cross-link data and would require further protein enhancement steps to increase cross-linking probabilities. In this paper, we show that a simple iBAQ analysis built from classical bottom-up data is a simple way to evaluate the chance of achieving good-quality XL-MS results for proteins of interest in complex samples. This approach can also be used to easily follow their enrichment.

As described below, we establish a protein abundance "cut-off" at 20% of the most abundant proteins, after which, in general, less than one cross-link per protein, and thus poor structural data, can only be expected. We therefore establish, for the first time, a rule, that is, the proteins of interest should be in the first quartile of abundant proteins (at the minimum) to ensure higher chances of attaining rich cross-linking data. This rule aims to provide some guidelines for those who are interested in using XL-MS to study a particular protein complex of interest and provide a starting point to ensure a successful cross-linking experiment.

EXPERIMENTAL SECTION

Products. NNP9 was synthesized by the COBRA Laboratory in Rouen, stored, and dissolved in dimethyl sulfoxide (DMSO) at 100 mM, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), ammonium bicarbonate, sodium deoxycholate (SDC), copper(II) sulfate, tris(3-

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513

hydroxypropyltriazolylmethyl)amine (THPTA), formic acid (FA), trifluoroacetic acid (TFA), and sodium ascorbate were purchased from SIGMA Aldrich. Photocleavable alkyne agarose beads (PCAB) were purchased from Click Chemistry Tools. The UV lamp (365 nm, 100 W) was purchased from Analytik Jena US. 96-well plates and Amicon Ultra 0.5 mL, 30 kDa were purchased from Dominique Dutscher.

***N. meningitidis* Whole-Cell Analysis (Figure 1A).** *N. meningitidis* 8013 (Nm8013) ΔsiaD mutant strains were spread on GCB plates supplemented with 5 μg/mL chloramphenicol, supplement I (D(+) glucose, L-glutamine, and cocarboxylase), and supplement II (ferric nitrate) and then grown overnight at 37 °C and 5% CO_2 in a Biosecurity II incubator (Figure 1A). Cells were harvested the following day, washed in PBS (Gibco, pH 7.4), and inactivated using 6 M urea.

No XL. The sample was reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP) for 30 min and then alkylated with 20 mM iodoacetamide for another 30 min. It was further diluted to reduce the urea concentration to less than 1 M in digestion buffer (50 mM ammonium bicarbonate). 10 μg of trypsin was added for overnight digestion at 37 °C. The digestion reaction was quenched the next day with 1% FA, and digests were precleaned on a C18 phase before LC−MS/MS analysis. 1 μg peptides were analyzed using a nanoLC system coupled to an Orbitrap Eclipse Tribrid (Thermo-Scientific) mass spectrometer and separated on a 50 cm C18 column with a 2 h linear gradient from 7 to 30% solvent B (80% acetonitrile and 0.1% FA) mixing with solvent A (0.1% FA), followed by a 30 min ramp up to 60% solvent B. The data were acquired using data-dependent mode with a 3 s cycle time. A survey scan was collected from 400 to 1700 in the Orbitrap (resolution: 60k at m/z 200) with a maximum injection time of 50 ms and an AGC (automatic gain control) target of 4×10^5. Dynamic exclusion was set to 25 s targeting charge stated at 2−7. Peptides were fragmented using HCD at 27% collision energy and acquired at a 30k resolution. Raw files were analyzed by using MaxQuant.

With XL. We used the raw files generated by Rey et al. in 2021[23] (PXD021553) that were obtained from a biological triplicate on an Orbitrap Q-Exactive HF mass spectrometer (Thermo-Scientific). For consistency, we reprocessed these raw files with the latest version of Mass Spec Studio.

***N. meningitidis* Membrane Pellet Analysis (Figure 1B).** Eight plates of Nm8013 ΔsiaD were prepared as described above, which was then harvested the following day and washed in PBS. The sample was centrifuged for 20 min at 3000 rpm and 4 °C, and the supernatant was carefully discarded. The pellet was resuspended and incubated with lysis buffer (50 mM HEPES, 300 mM NaCl, 100 μg/mL lysozyme, 5% glycerol, 2.5 μL Benzonase Nuclease, 0.1 mM EDTA, and 1 mM $MgCl_2$, pH 8.0) for 3 h at 4 °C with a rotational shaker, then followed by two steps of sonication for 2 min each at 4 °C, 10 s intervals, and 30% amplitude. A low-speed centrifugation was first performed at 4000 rpm for 30 min to remove unbroken cells and debris. The supernatant was then subjected to an ultracentrifugation step at 100,000g for 45 min at 4 °C using a type 70Ti fixed-angle rotor (Beckman Coulter). The supernatant containing the soluble protein fraction was removed and stored at −20 °C, while the membrane pellets were washed three times with buffer (50 mM HEPES, pH 8.0) and left in buffer overnight to achieve a soft pellet.

No XL. The membrane pellet was resuspended in 8 M urea, and the supernatant (500 μg) was then buffer-exchanged to 8 M urea. The samples were then reduced and alkylated, as previously stated. For digestion, the supernatant sample was diluted in a digestion buffer to achieve less than 1 M urea. The membrane pellet was buffer-exchanged in 0.5 mL of 10 kDa Amicon with digestion buffer, and SDC was added to a final concentration of 1% to emulate digestion conditions of the cross-linking experiment. Tryptic digestion and LC−MS/MS for both the membrane pellet and supernatant were carried out as for the whole cell described above.

With XL. The membrane pellet was resuspended in 0.5 mL of fresh buffer, then 60 μL of 10 mM NNP9 in DMSO was added and allowed to react for 3 h at room temperature. The cross-linked membrane pellet was transferred to 0.5 mL of 10 kDa Amicon. The excess cross-linker was removed by 6 concentration−dilution cycles and buffer-exchanged with digestion buffer. The membrane pellet was removed from the filter and placed into a 1.5 mL Protein LoBind Eppendorf tube where the digestion buffer was topped to 500 μL; SDC was added to 1% to aid in membrane protein digestion, and 5 μg of trypsin was added for overnight digestion at 37 °C. The reaction was acidified with 2% TFA for 10 min, and the supernatant was removed. For click chemistry enrichment, PCAB (1 μL:20 nmol) were mixed with tryptic peptides, copper(II) sulfate (5:1), THPTA (25:1), and sodium ascorbate (50:1), where ratios are based on NNP9 quantity. After 90 min of shaking (1000 rpm, room temperature), the supernatant was removed, and 5 mM TCEP was added for 30 min, followed by 20 mM iodoacetamide for 30 min. Beads were then transferred to a 200 μL tip blocked with filter paper (10 μL PCAB per tip) and washed with acetonitrile/isopropanol/0.1% FA (1:1:1) 6 times in a Stage Tip centrifuge, followed by 3 times with 0.1% FA. PCAB were transferred to a 96-well plate (maximum 10 μL of PCAB per well), covered with 100 μL of 0.1% FA, and subject to UV and rotational mixing (700 rpm) for 1 h. Released peptides were pipetted and transferred to the injection vial. Eluted peptides from PCAB were analyzed by nanoLC-MS/MS using a Vanquish Neo UHPLC system (Thermo Scientific) coupled to an Orbitrap Eclipse Tribrid mass spectrometer fitted with an EASY-Spray Source (Thermo Scientific). Peptides were separated on an EASY-Spray PepMap Neo (75 μm × 500 mm, 2 μm particle size, and 100 Å pore size, Thermo Scientific, San Jose, CA), heated at 60 °C with a 3 h linear gradient from 8 to 30% solvent B, followed by a 35 min ramp up to 60% solvent B. The data were acquired in data-dependent mode with a 3 s cycle time. A survey scan was collected from 300 to 1500 in the Orbitrap (resolution: 60k at m/z 200) with a maximum injection time of 100 ms and an AGC target of 4×10^5. Dynamic exclusion was set to 30 s targeting charge stated at 3−8. Peptides were fragmented using HCD at 27% collision energy, and MS/MS spectra were acquired at a 30k resolution.

Identification of Cross-Linked Peptides. All raw files were analyzed using Mass Spec Studio[32] (v2.4) CRIMP 2.0,[33] including those from Rey et al. 2021[23] for consistency. Processed files were exported as spectrum matches. As Mass Spec Studio controls FDR at the level of the selected export, FDR was controlled at the spectrum match level, and the files were subsequently processed using in-house scripts to accept all identifications within a 1% FDR, remove duplicates within and between replicates, and assign XL binding sites. From here on, the total number of cross-links is referred to as the unique number of cross-links found in the combined replicate data set. PPI networks were created using Cytoscape (v3.9.1). The

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513

number of cross-links per protein was calculated by taking a unique list of all proteins found (Protein 1 and Protein 2) when filtering labeled peptides with the type: "cross-linked" and counting the frequency of their appearance in the cross-link data set. Whole cell and membrane cross-link data can be found in Table S1.

iBAQ and Average Cross-Link Analysis. Proteins were ranked by decreasing the order of abundance calculated by the iBAQ value from MaxQuant. The iBAQ was calculated by dividing the total sum of all peptide peak intensities by the total number of theoretical observable peptides for a given protein.[34] Their percent abundance was calculated by ranking from the total proteins identified and grouped into 5% protein abundance intervals. Cross-linked proteins and their calculated cross-linked frequencies were matched with those found in the bottom-up data. If a protein was not found to be cross-linked, then the cross-link frequency was assigned to 0. An average per protein group was then calculated and plotted alongside decreasing iBAQ values.

Data Availability. All raw data, parameter files, and processed data have been deposited to the ProteomeXchange Consortium via the PRIDE[35] partner repository with the data set identifier PXD045792.

RESULTS AND DISCUSSION

Analyzing the Depth of an XL-MS Experiment. To start, we performed a bottom-up proteomics analysis of *N. meningitidis* whole cells and obtained relative protein abundances through the iBAQ method. Figure 2 shows all

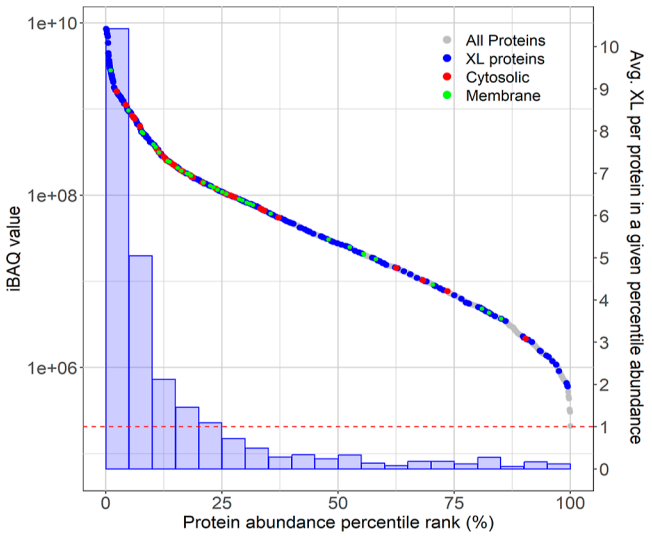


Figure 2. Average cross-link per protein abundance percentile. iBAQ values for all proteins found in a bottom-up analysis of *N. meningitidis* (gray) with proteins found cross-linked in an in vivo experiment (blue). The bar plots represent the average cross-link per protein calculated per 5% protein abundance percentile with a cut off at 1 cross-link per protein, highlighted as a dashed red line.

the proteins plotted from most to least abundant (gray) according to their iBAQ rankings and the coverage of cross-linked proteins from the in vivo experiment performed by Rey et al. (blue). The densest region of cross-linked proteins arises from the proteins that are at least in the top 50% abundance range, yet quite a few proteins were detected in the last quartile showing that indeed an in vivo XL-MS experiment can cover a large dynamic range (4 log-fold). We then went a step further to see how cross-link numbers correlated with abundance levels. We calculated the average number of cross-links per 5% protein abundance intervals, which is shown as a bar plot. The horizontal red dashed line represents an average of 1 cross-link per protein. This threshold is quickly reached after the first top 20% abundant proteins, with a steep decrease from the first 5% of abundant proteins to the next interval as the average is halved. This drastic decrease of average cross-links in the first top 10% of proteins highlights the essential dependence of protein abundance on cross-linking. Proteins that were found to be cross-linked below the top 20th percentile of abundance also showed less reproducibility between replicates (Figure S1). This indicates that if proteins of interest are below this abundance range, cross-link data can be expected to be produced more randomly.

We then searched to see where some well-known and relatively abundant membrane complexes rank, along with some well-established cytosolic proteins involved in metabolism and biosynthesis. For the membrane, we chose three complexes that cover the inner membrane (ATP synthase, catalyzes the formation of ATP from ADP), spanning both inner and outer membranes (Pil proteins from the type IV piliation system virulence factor), and the outer membrane [BAM (β-barrel assembly machine) complex folds and inserts β-barrel proteins in the outer membrane]. For the cytoplasm, we selected proteins from acetyl-CoA carboxylase complex (acc),[36] arginine biosynthesis proteins (arg),[37] purine biosynthesis (pur),[38] and proteins involved in the tricarboxylic acid cycle (suc)[39] as these were rather abundant groups of proteins involved in important metabolic pathways. Their relative abundance positions are highlighted in green (membrane) and red (cytoplasm) in Figure 2. In the first 10%, only a few membrane proteins appear (mostly near the 10% abundance border), while the cytosolic proteins appear to have a better distribution in this range. Furthermore, 21 of the 37 membrane proteins are below the top 20% range, with a larger spread throughout the curve. While 12 of the 28 cytosolic proteins are below 20%, yet still concentrated toward the higher abundance range. With membrane proteins beginning to appear mainly after the top 10% of most abundant proteins and having to compete with cytosolic proteins, it is no surprise why a majority of the cross-link data were on cytoplasmic proteins.

Overall, for proteins not in the 20th percentile, which is the case for most membrane proteins in our data, the chance to achieve more than one cross-link, even with a cross-link enrichment strategy that is as efficient as ours, is very low. In general, targeting membrane proteins may thus not be optimal with a whole-cell experiment, and an additional enrichment step is required to push them closer to the abundance "zone" where useful XL-MS can be expected. To test this hypothesis, we decided to enrich *N. meningitidis* membrane proteins and assess if our 20% abundance cutoff rule was still effective and thus boost the number of identified membrane protein cross-links.

Enhancing Cross-Linking Information on the Membrane Proteome. We tested three different methods before the cross-linking step: (1) Detergent extraction—various detergents were tested to aid in membrane protein digestion and therefore obtain more cross-link data. (2) Ghost cells—cells were treated with antibiotics to create holes in the membrane allowing cytosolic components to leak out, leaving mostly membrane. (3) Isolating the membrane pellet. The first

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513

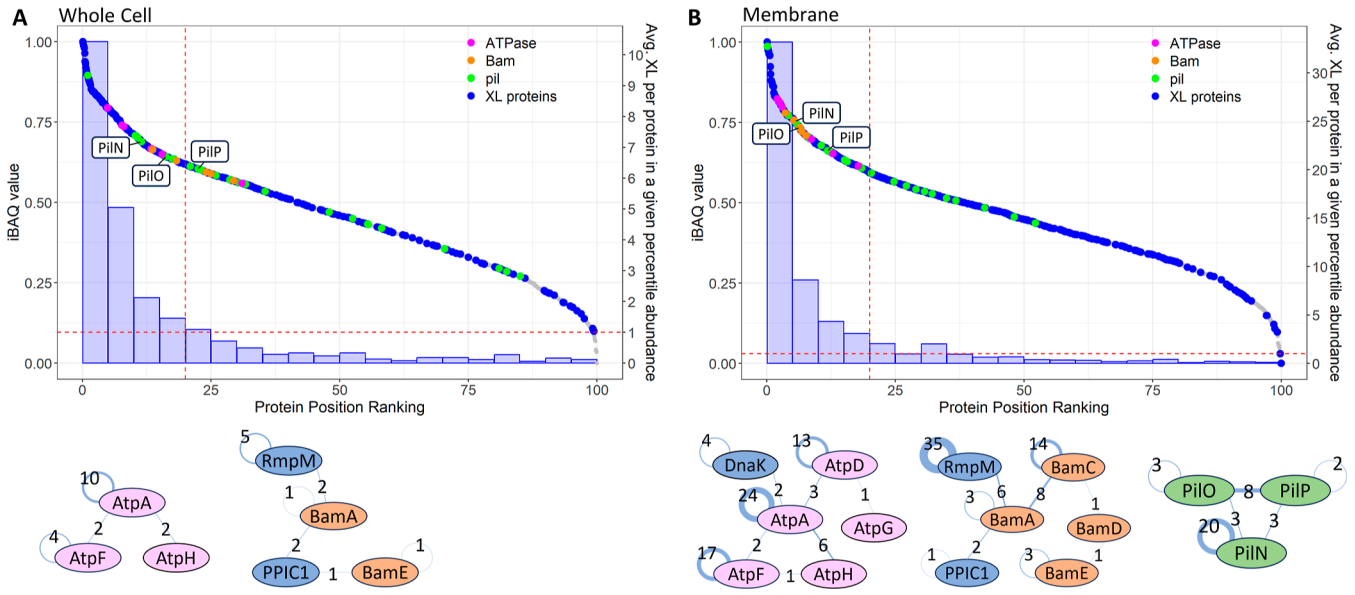


Figure 3. Comparison of three membrane assembly abundances and cross-linking information obtained on their networks. Proteins from ATPase (pink), BAM (orange), and T4P (green) are highlighted in both the whole cell (A) and membrane pellet (B) conditions. The networks on these three assemblies obtained from cross-linking experiments are given to the right of the graphs. Average cross-link per 5% abundance percentile range is given as a bar graph for the membrane pellet cross-link data set (A) and coverage of cross-linked proteins is shown in blue.

two strategies did not lead to successful enrichment of membrane proteins, and no difference was observed in cross-link numbers for cytosolic and membrane proteins compared with the whole-cell control. However, classic membrane pellet isolation was found to be a simple but efficient enrichment method that led to interesting results.

To evaluate whether our enrichment is sufficient for us to continue with a cross-linking experiment, we followed the normalized iBAQ and percent ranking values for the same membrane and cytosolic proteins mentioned above. All three membrane protein complexes showed both higher iBAQ values and a shift to a higher percent ranking (Table S2). The strictly outer membrane BAM complex showed the most enrichment, while the inner membrane ATPase complex showed the least, yet still nearly halving its top percent ranking position (top 13 to 7% abundance). The cytoplasmic protein complexes decreased their position ranking in the membrane pellet by a factor of 1.5 to 3.5 compared to the whole-cell control condition, showing a decent shift toward lower abundances in the membrane pellet condition. With a well-enhanced membrane proteome, we repeated a triplicate cross-linking experiment this time on the membrane pellets to see if it changed the number and nature of the cross-links seen and how much further into the membrane proteome we could probe.

As in Figure 2, we used our cross-linking data set from the membrane pellet to recalculate the average cross-links per abundance percentile (Figure 3B). In blue are all the proteins found cross-linked in our membrane digest, showing good coverage of cross-linking data on the range of proteins present in the membrane pellet. Plotting the average cross-links obtained per 5% abundance increments, we again see that most of the cross-links occur in the top 5% of proteins with a drastic decrease in the following interval ranges. Interestingly, the point at which an average of less than one cross-link per protein is reached is maintained around 20%. The minimum abundance range here is pushed slightly to 25% but only with a mere increase of an average of 2 cross-links per protein despite almost 4× more identifications. Reproducibility of the cross-links was again checked between those above and below the 20% cutoff (Figure S2), and as with the whole-cell data, higher reproducibility was seen for the top 20% of cross-linked proteins. The most striking change occurred within the top 20% abundances, where there was quite a significant increase in the average cross-link numbers compared to the whole-cell study. This can possibly be due to not having reached a saturation in cross-link identifications of the topmost abundant proteins. Optimization of various factors, such as the cross-linker to protein ratio, fragmentation efficiency, and identifications in data processing, may be far from being reached in current XL-MS experiments. Improvements in the XL-MS method will first lead to an increase in the number of cross-links in high-abundance protein regions, rather than enhancing cross-link information across all abundance ranges.

Probing Membrane-Associated Protein Networks. Proteins from each of the three selected membrane networks—ATPase subunits (pink), BAM complex (orange), and Pil proteins (green)—are highlighted for both the whole cell (Figure 3A) and the membrane pellet (Figure 3B). To see whether their overall increase in abundance resulted in better cross-linking data, we analyzed the networks we were able to obtain in each data set. All interactions between proteins in the same group (same color) were kept, but a filtering of at least 2 different cross-links or higher was done for proteins not in the same group to ensure the removal of any false positive interactions. In the whole-cell data, as the ATPase proteins were mostly above the cutoff, we were unsurprisingly able to obtain sufficient cross-link information to establish a small network. Searching for BAM proteins in the whole cell data set, a small network was found, but no interaction between the BAM proteins was detected. This is correlated to the abundance of BAM proteins tethering the threshold cutoff. A search of Pil proteins resulted in mostly intra-XLs, resulting in an absence of a network. In the membrane pellet data set, a

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513

more complete network between the proteins from the ATPase and the BAM complex was established along with reinforcement of most of the cross-links found in the whole-cell experiment. For the ATPase, the interactions of AtpA with AtpF and AtpH were reinforced with more cross-links found between the three proteins, and a more thorough network established by capturing the interaction with AtpD and AtpG. Furthermore, the protein-folding chaperone DnaK has a high ATP affinity, which can explain its association with the ATPase.[40] For the BAM complex, the full network of BAM proteins was seen in the membrane cross-linking condition as well as reinforcement of the interaction of BamA with RmpM, which was shown to be part of the BAM complex in *Neisseria* to help protein stability rather than folding.[41] Peptidyl-prolyl cis/trans isomerases (PPIases) assist in protein folding by catalyzing the *cis/trans* isomerization of peptide bonds before the formation of a proline. They have been implicated in bacterial virulence by aiding in the correct folding of outer membrane proteins involved in bacterial virulence such as adhesins and have been shown to work together with the BAM complex in *Escherichia coli*.[42] We were additionally able to capture cross-links between PilN, PilO, and PilP, which are proteins from the alignment complex in the inner membrane and periplasm, which we have not been able to identify prior to this.

To ensure that this enhanced cross-link information was due to enrichment of membrane proteins toward the top 20% abundance range and not just an overall increase in cross-link numbers due to the lower MS^2 resolution, we show the total contribution of the cross-links from the three membrane assemblies compared to the total cross-links in the data set (Table 1). The total percent contribution of cross-links from each of the three membrane complexes is higher in the membrane pellet condition than in the whole cell. The highest percent increase is seen for the Pil proteins. The smaller percentage increase of cross-links in the ATPase and BAM complex proteins correlates to their initial abundances in the whole-cell condition already being quite high. Six out of the 7 ATPase protein subunits were already in the top 20%, as well as 2 out of the 5 BAM complex proteins. In the membrane pellet, all ATPase and BAM complex proteins were among the top 20%. A much larger difference in the Pil protein cross-link percentage is seen as their total count in the top 20% from the whole cell to the membrane pellet went from 8 to 15. Additionally, from some of our optimization experiments, we had a duplicate data set of membrane pellet cross-linking acquired at 60k MS^2 resolution where we obtained about half the cross-links compared to the presented membrane pellet data, yet the percent contribution from the three membrane assembly proteins was equivalent (Table S4).

Table 1. Number of Unique Cross-Links and Cross-Linked Proteins in the Whole Cell and Membrane Pellet Conditions with the Total Cross-Link Contribution from the ATPase, BAM Complex, and Pil Proteins

label type	whole cell	membrane
total unique XLs	**1612**	**4270**
BAM	10	75
ATPase	39	160
Pil	60	434
total XL proteins	**327**	**494**

Assessing the Role of Databases in Searches. As mentioned in the previous section, many XL-MS studies have used reduced protein databases for their searches. To see if that alone would have an improvement on identified XLs, we decided to reprocess the whole-cell XL data set with restricted databases. Three restrictions were chosen: (1) taking only dead-ends, as suggested by Lima and colleagues[43] (DE restricted), (2) proteins were only found in the iBAQ curve (iBAQ restricted), and (3) proteins in the top 20% (top 20 restricted). The results are shown in Table S5. With improvements in the last version of Mass Spec Studio CRIMP search algorithms, there was no noticeable gain in time or labeled peptide information in both the DE- and iBAQ-restricted databases. Interestingly, there was an increase in overall XL numbers seen with the top 20 restricted database. Comparing the XLs in the three membrane complexes to the total protein database, the top 20 restricted database had the highest gain in Pil protein and ATPase XLs (9 and 6, respectively) but with a drastic decrease in BAM complex XLs. The decrease in BAM XLs is evident as only 2 of the 5 proteins were in the top 20% but the other 3 were still rather high in abundance to be well identified. A closer look at the gain of ATP cross-links shows that 6 single-hit cross-links between an ATP subunit and non-ATP proteins were added. From the 9 XLs gained on Pil proteins, 3 were intra-PilQ, with one showing low scores in a single replicate, and 3 were between PilQ and TsaP, with 2 showing low scores and 1 found in a single replicate. The remaining three cross-links were single-hit inter-XLs with non-Pil proteins. With improvements in search algorithms, there seems to be no beneficial gain in restricting the database. Moreover, restricting too much can lead to unfavorable results. Overall, simply reducing the database did not result in increased XL sensitivity, thus highlighting the importance of protein enhancement prior to cross-linking and not adjusting for abundant proteins post-cross-linking.

These analyses on our data confirm the rule of establishing a protein abundance cut off at 20% to achieve valuable cross-link information. To assess whether this quick quality control test holds true for other data sets, we decided to evaluate large-scale publicly available cross-linking data sets for which we already had standard bottom-up proteomics data.

Evaluating Our Quality Control Test in Other Data Sets. To see if this trend is observed in other XL-MS experiments, three data sets were selected on organisms with various complexities: *E. coli*[44] (reference proteome: 4403), *Saccharomyces cerevisiae*[45] (reference proteome: 6060), and HeLa[44] (reference proteome: 82,678). Each data set was conducted with at least two different cross-linker concentrations, which allowed us to evaluate whether this could influence the abundance cutoff. iBAQ curves showing cross-link protein coverage and calculated cross-link averages are shown in Figure S3. Cross-linking on all three organisms was performed using DSSO, a cleavable nonenrichable cross-linker with NHS ester groups. In comparison, NNP9 is noncleavable, enrichable, and contains NHS carbamate groups, which have shown differences in the rate of hydrolysis.[30] Cross-link data was obtained through extensive fractionation in all three cases (6 to 22 fractions) and some in duplicate. Despite differences in the cross-linkers and the concentrations used, similar trends in average cross-links per protein abundance range are seen in all data sets. The top 5% of proteins contain most of the cross-link information followed by a very steep drop in subsequent percentile ranges, where the average of less than one XL per

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513

protein was reached between 15 and 25% of the top abundant proteins. Furthermore, comparing with our data, the total average of cross-links even in the first 5% abundance range is lower in all three data sets. These data indicate that should a system be studied with a nonenrichable cross-linker, proteins of interest should be closer to the top 10−15% of abundance for well-rounded XL-MS data as the overall cross-link hits will be lower on the whole data set.

Of course, increasing the cross-linker/protein ratio, adding fractionation steps, and relaxing the FDR used for the search can improve the overall number of cross-link identifications. Thus, it is not necessary that an average of less than one cross-link is reached after 20% or even across the entire protein abundance range. Although useful cross-link information can still be found after our proposed cutoff, the main purpose of our analysis was to show that most of the cross-links obtained in an XL-MS experiment are found for the top 20% most abundant proteins and overshadow identification on proteins of lower abundance ranges.

There is still lots of room for improvement for getting more sensitive and robust XL-MS. Such improvements can involve a more effective way for removing dead-ends and thus enriching in only cross-links, additional ion mobility for orthogonal separation, and moving toward DIA experiments for more robust identifications. With all of these advances, we hope to be able to lower the "top 20%" threshold. A new sensitivity threshold can easily be re-evaluated with our proposed method.

CONCLUSIONS

Although it is known that cross-linking results are typically dominated by higher abundance proteins, here, we established an abundance threshold below which little to no valuable XL data can be expected. This leads to capturing more complete protein interaction network data and intraprotein cross-links for structural information. We have demonstrated that this threshold lies in the first top 20% of abundant proteins, and even within this rather narrow abundance range, there is a drastic decrease of cross-links that can be expected of proteins in the first and second half. Testing this method on several data sets, we show that despite the complexity of the organism or the amount and type of cross-linker used, it is still the top 20% abundant proteins that show the most cross-links. Furthermore, cross-links of proteins below this cutoff are not only scarcer but less reproducible, resulting in a more random chance that proteins of interest in the lower abundance ranges will get cross-linked. We hope that this general rule based on the use of a simple iBAQ quantification will serve as a guideline for those interested in studying a complex of interest with XL-MS, especially in complex samples, and ensure conditions for a successful cross-linking experiment from the get-go.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c04682.

Venn diagrams comparing cross-link reproducibility between replicates of the whole cell and the membrane pellet; average XL for protein abundance; comparison of abundances of various membranes and cytosolic proteins in three different digests; total proteins identified in all digests; cross-links found in a duplicate analysis of the membrane pellet acquired at 60k MS^2; and labeled peptide identification using variously restricted databases (PDF)

List of cross-links found in the whole cell and membrane pellet cross-link experiments and with restricted FASTAs (XLSX)

AUTHOR INFORMATION

Corresponding Author

Julia Chamot-Rooke − *Institut Pasteur, Université Paris Cité, CNRS UAR 2024, Mass Spectrometry for Biology Unit, Paris 75015, France;* orcid.org/0000-0002-9427-543X; Email: julia.chamot-rooke@pasteur.fr

Authors

Lucienne Nouchikian − *Institut Pasteur, Université Paris Cité, CNRS UAR 2024, Mass Spectrometry for Biology Unit, Paris 75015, France;* orcid.org/0009-0003-4991-6159

David Fernandez-Martinez − *Institut Pasteur, Université Paris Cité, INSERM UMR1225, Pathogenesis of Vascular Infections Unit, Paris 75015, France*

Pierre-Yves Renard − *Univ Rouen Normandie, INSA Rouen Normandie, CNRS, Normandie Univ, COBRA UMR 6014, INC3M FR 3038, Rouen F-76000, France;* orcid.org/0000-0001-9094-9778

Cyrille Sabot − *Univ Rouen Normandie, INSA Rouen Normandie, CNRS, Normandie Univ, COBRA UMR 6014, INC3M FR 3038, Rouen F-76000, France;* orcid.org/0000-0002-5014-7217

Guillaume Duménil − *Institut Pasteur, Université Paris Cité, INSERM UMR1225, Pathogenesis of Vascular Infections Unit, Paris 75015, France*

Martial Rey − *Institut Pasteur, Université Paris Cité, CNRS UAR 2024, Mass Spectrometry for Biology Unit, Paris 75015, France;* orcid.org/0000-0002-7378-1106

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c04682

Author Contributions

All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the European Union's Horizon 2020 Research and Innovation Program under grant nos. 823839 (EPIC-XS), and the French Research Agency (ANR-18-CE11-0022 and T4PNanoAction project). J.C.R., L.N., and M.R. acknowledge the Region Ile de France (DIM 1HEALTH) and the LABEX IBEID (grant no. ANR-10-LABX-62-IBEID from the Programme d'Investissements d'Avenir) for funding the Eclipse Orbitrap. Graphical abstract and Figure 1 were created with BioRender.com.

REFERENCES

(1) Tang, X.; Wippel, H. H.; Chavez, J. D.; Bruce, J. E. *Protein Sci.* **2021**, *30*, 773−784.
(2) Graziadei, A.; Rappsilber, J. *Structure* **2022**, *30*, 37−54.
(3) Chavez, J. D.; Bruce, J. E. *Curr. Opin. Chem. Biol.* **2019**, *48*, 8−18.
(4) Lee, K.; O'Reilly, F. J. *Essays Biochem.* **2023**, *67*, 215−228.

(5) O'Reilly, F. J.; Rappsilber, J. *Nat. Struct. Mol. Biol.* **2018**, *25*, 1000−1008.
(6) Sinz, A. *Cell Syst.* **2018**, *6*, 10−12.
(7) Britt, H. M.; Cragnolini, T.; Thalassinos, K. *Chem. Rev.* **2022**, *122*, 7952−7986.
(8) Rappsilber, J. *J. Struct. Biol.* **2011**, *173*, 530−540.
(9) Chen, Z.; Fischer, L.; Tahir, S.; Bukowski-Wills, J.-C.; Barlow, P.; Rappsilber, J. *Wellcome Open Res.* **2016**, *1*, 5.
(10) Schmidt, C.; Urlaub, H. *Curr. Opin. Struct. Biol.* **2017**, *46*, 157−168.
(11) Greber, B. J.; Boehringer, D.; Leitner, A.; Bieri, P.; Voigts-Hoffmann, F.; Erzberger, J. P.; Leibundgut, M.; Aebersold, R.; Ban, N. *Nature* **2014**, *505*, 515−519.
(12) Ziemianowicz, D. S.; Kosinski, J. *Curr. Opin. Struct. Biol.* **2022**, *77*, 102488.
(13) Schneider, M.; Belsom, A.; Rappsilber, J. *Trends Biochem. Sci.* **2018**, *43*, 157−169.
(14) Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. *Chem. Rev.* **2022**, *122*, 7500−7531.
(15) Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A. *Protein Sci.* **2018**, *27*, 245−258.
(16) Filella-Merce, I.; Bardiaux, B.; Nilges, M.; Bouvier, G. *Structure* **2020**, *28*, 75−82.e4.
(17) Stahl, K.; Graziadei, A.; Dau, T.; Brock, O.; Rappsilber, J. *Nat. Biotechnol.* **2023**, *41*, 1810−1819.
(18) Lenz, S.; Sinn, L. R.; O'Reilly, F. J.; Fischer, L.; Wegner, F.; Rappsilber, J. *Nat. Commun.* **2021**, *12*, 3564.
(19) Matzinger, M.; Mechtler, K. *J. Proteome Res.* **2021**, *20*, 78−93.
(20) Ruwolt, M.; Piazza, I.; Liu, F. *Curr. Opin. Struct. Biol.* **2023**, *82*, 102648.
(21) Tang, X.; Bruce, J. E. *Mol. Biosyst.* **2010**, *6*, 939−947.
(22) Mendes, M. L.; Fischer, L.; Chen, Z. A.; Barbon, M.; O'Reilly, F. J.; Giese, S. H.; Bohlke-Schneider, M.; Belsom, A.; Dau, T.; Combe, C. W.; Graham, M.; Eisele, M. R.; Baumeister, W.; Speck, C.; Rappsilber, J. *Mol. Syst. Biol.* **2019**, *15*, No. e8994.
(23) Rey, M.; Dhenin, J.; Kong, Y.; Nouchikian, L.; Filella, I.; Duchateau, M.; Dupré, M.; Pellarin, R.; Duménil, G.; Chamot-Rooke, J. *Anal. Chem.* **2021**, *93*, 4166−4174.
(24) Gao, H.; Zhao, L.; Zhong, B.; Zhang, B.; Gong, Z.; Zhao, B.; Liu, Y.; Zhao, Q.; Zhang, L.; Zhang, Y. *Anal. Chem.* **2022**, *94*, 7551−7558.
(25) de Jong, L.; de Koning, E. A.; Roseboom, W.; Buncherd, H.; Wanner, M. J.; Dapic, I.; Jansen, P. J.; van Maarseveen, J. H.; Corthals, G. L.; Lewis, P. J.; Hamoen, L. W.; de Koster, C. G. *J. Proteome Res.* **2017**, *16*, 2457−2471.
(26) Steigenberger, B.; Pieters, R. J.; Heck, A. J. R.; Scheltema, R. A. *ACS Cent. Sci.* **2019**, *5*, 1514−1522.
(27) Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F.-U.; Ban, N.; Malmström, L.; Aebersold, R. *Science* **2012**, *337*, 1348−1352.
(28) Sinn, L. R.; Giese, S. H.; Stuiver, M.; Rappsilber, J. *Anal. Chem.* **2022**, *94*, 4627−4634.
(29) Wittig, S.; Ganzella, M.; Barth, M.; Kostmann, S.; Riedel, D.; Pérez-Lara, A.; Jahn, R.; Schmidt, C. *Nat. Commun.* **2021**, *12*, 858.
(30) Nury, C.; Redeker, V.; Dautrey, S.; Romieu, A.; van der Rest, G.; Renard, P.-Y.; Melki, R.; Chamot-Rooke, J. *Anal. Chem.* **2015**, *87*, 1853−1860.
(31) Rouphael, N. G.; Stephens, D. S. *Methods Mol. Biol.* **2012**, *799*, 1−20.
(32) Rey, M.; Sarpe, V.; Burns, K.; Buse, J.; Baker, C. A. H.; van Dijk, M.; Wordeman, L.; Bonvin, A. M. J. J.; Schriemer, D. C. *Structure* **2014**, *22*, 1538−1548.
(33) Crowder, D. A.; Sarpe, V.; Amaral, B. C.; Brodie, N. I.; Michael, A. R. M.; Schriemer, D. C. *Anal. Chem.* **2023**, *95*, 6425−6432.
(34) Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. *Nature* **2011**, *473*, 337−342.
(35) Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, J. *Nucleic Acids Res.* **2022**, *50*, D543−D552.
(36) Tong, L. *Cell. Mol. Life Sci.* **2005**, *62*, 1784−1803.
(37) Hani, E. K.; Ng, D.; Chan, V.-L. *Can. J. Microbiol.* **1999**, *45*, 959−969.
(38) Zhang, Y.; Morar, M.; Ealick, S. E. *Cell. Mol. Life Sci.* **2008**, *65*, 3699−3724.
(39) Schoen, C.; Kischkies, L.; Elias, J.; Ampattu, B. J. *Front. Cell. Infect. Microbiol.* **2014**, *4*, 114.
(40) Wawrzynów, A.; Zylicz, M. *J. Biol. Chem.* **1995**, *270*, 19300−19306.
(41) Volokhina, E. B.; Beckers, F.; Tommassen, J.; Bos, M. P. *J. Bacteriol.* **2009**, *191*, 7074−7085.
(42) Ünal, C. M.; Steinert, M. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 544−571.
(43) Lima, D. B.; De Lima, T. B.; Balbuena, T. S.; Neves-Ferreira, A. G. C.; Barbosa, V. C.; Gozzo, F. C.; Carvalho, P. C. *J. Proteomics* **2015**, *129*, 51−55.
(44) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. *Nat. Methods* **2015**, *12*, 1179−1184.
(45) Fürsch, J.; Kammer, K.-M.; Kreft, S. G.; Beck, M.; Stengel, F. *Anal. Chem.* **2020**, *92*, 4016−4022.

https://doi.org/10.1021/acs.analchem.3c04682
Anal. Chem. 2024, 96, 2506−2513