analytical chemistry

pubs.acs.org/ac Article

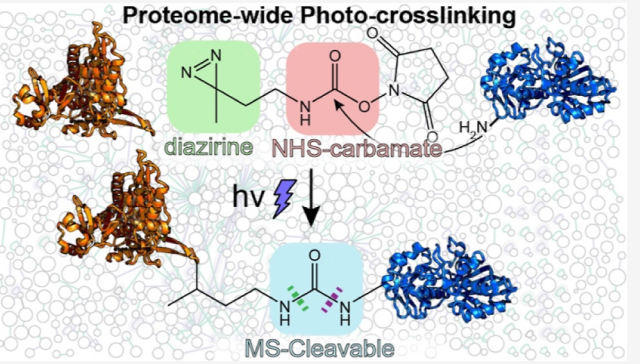
Progress toward Proteome-Wide Photo-Cross-Linking to Enable Residue-Level Visualization of Protein Structures and Networks In Vivo

Anneliese M. Faustino,[§] Piyoosh Sharma,[§] Edgar Manriquez-Sandoval, Divya Yadav, and Stephen D. Fried*

Cite This: *Anal. Chem.* 2023, 95, 10670−10685 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cross-linking mass spectrometry (XL-MS) is emerging as a method at the crossroads of structural and cellular biology, uniquely capable of identifying protein−protein interactions with residue-level resolution and on the proteome-wide scale. With the development of cross-linkers that can form linkages inside cells and easily cleave during fragmentation on the mass spectrometer (MS-cleavable cross-links), it has become increasingly facile to identify contacts between any two proteins in complex samples, including in live cells or tissues. Photo-cross-linkers possess the advantages of high temporal resolution and high reactivity, thereby engaging all residue-types (rather than just lysine); nevertheless, photo-cross-linkers have not enjoyed widespread use and are yet to be employed for proteome-wide studies because their products are challenging to identify. Here, we demonstrate the synthesis and application of two heterobifunctional photo-cross-linkers that feature diazirines and *N*-hydroxy-succinimidyl carbamate groups, the latter of which unveil doubly fissile MS-cleavable linkages upon acyl transfer to protein targets. Moreover, these cross-linkers demonstrate high water-solubility and cell-permeability. Using these compounds, we demonstrate the feasibility of proteome-wide photo-cross-linking *in cellulo*. These studies elucidate a small portion of *Escherichia coli*'s interaction network, albeit with residue-level resolution. With further optimization, these methods will enable the detection of protein quinary interaction networks in their native environment at residue-level resolution, and we expect that they will prove useful toward the effort to explore the molecular sociology of the cell.



INTRODUCTION

With recent technical and technological advances, cross-linking mass spectrometry (XL-MS) is emerging as a powerful structural proteomics technique—uniquely capable of elucidating protein−protein interaction (PPI) networks at residue-level resolution and on the proteome-wide scale.[1−4] Specifically, improvements in high-resolution mass analyzers, chemical cross-linkers, and analysis software have joined forces to move XL-MS from a structural biology method (applied to purified protein complexes) to an "-omics" technique that bridges between cellular and structural biology. For example, in a recent report, O'Reilly and co-workers combined in-cell XL-MS with cryo-electron tomography (cryo-ET) to develop a structural model of the RNA polymerase (RNAP)−ribosome complex performing coupled transcription-translation in situ.[5]

Early cross-linkers applied in XL-MS studies, such as disuccinimidyl suberate and bis(sulfosuccinimidyl)suberate (BS3), are homobifunctional compounds consisting of two NHS-esters connected by a hydrocarbon linker.[6−11] Following chemical cross-linking and trypsinolysis, these compounds generate pairs of peptides (called cross-linked peptides) irreversibly adjoined between two nucleophilic residues (typically lysines). These cross-linkers proved useful for obtaining high coverage on single proteins; however, they were challenging to apply to complex samples, such as lysates, cells, or tissues. This is because the complexity of the search space associated with possible cross-linked peptides scales as the number of linear peptides squared (known as the N^2 problem),[12] which complicates spectral identification and raises the chances of false discovery.

Proteome-wide XL-MS studies emerged with the onset of a newer generation of cross-linkers,[13,14] such as disuccinimidyl sulfoxide and disuccinimidyl dibutyric urea (DSBU, as shown in Figure 1A), which can be cleaved in the gas phase during

Received: March 29, 2023
Accepted: June 9, 2023
Published: June 21, 2023

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

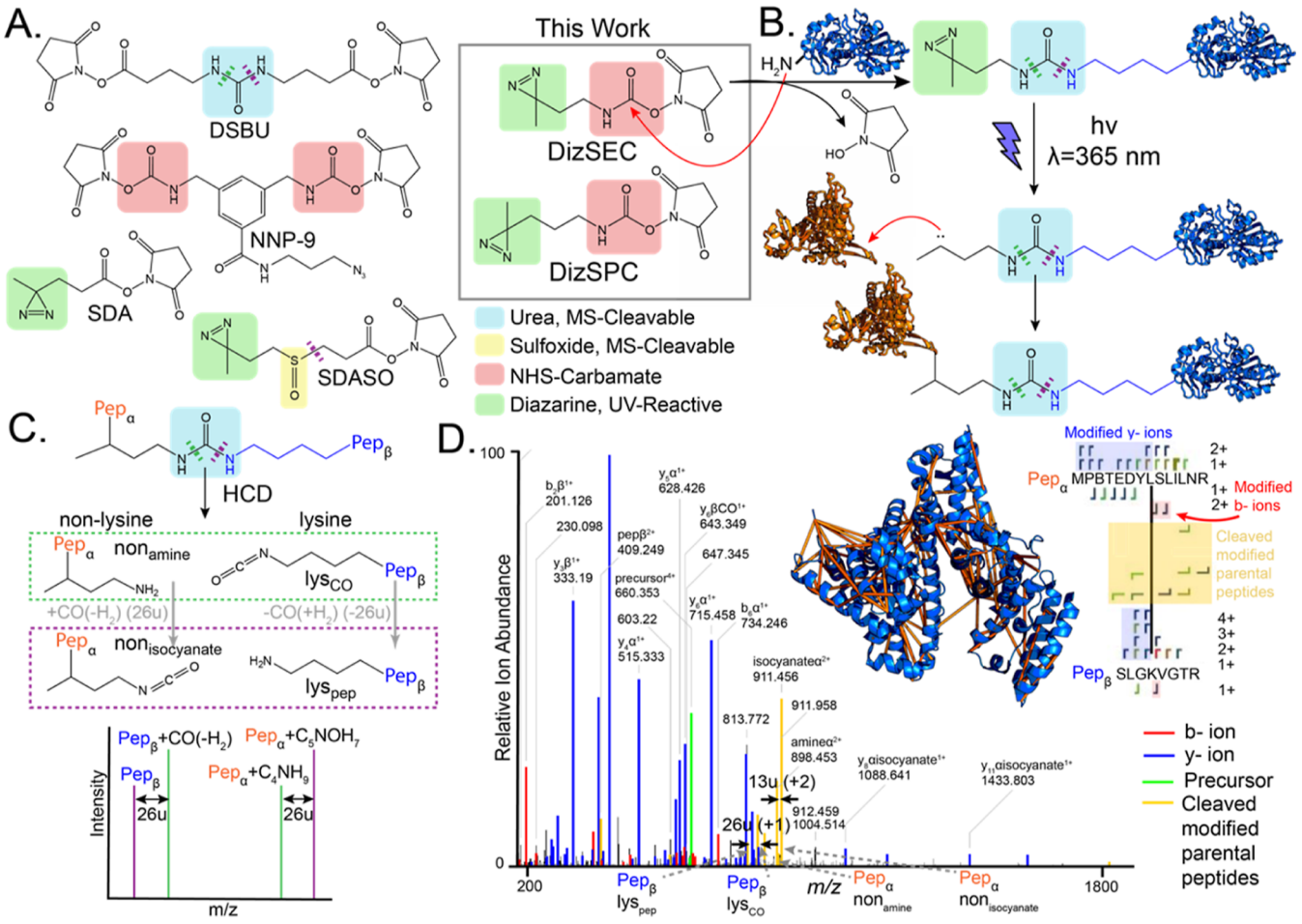


Figure 1. Structure and properties of MS-cleavable heterobifunctional photo-cross-linkers. (A) Structures of DSBU, a urea-based (teal box) cleavable cross-linker; NNP-9, an NHS-carbamate (red box) containing-cross-linker; SDA, a heterobifunctional diazirine (green box) cross-linker; and SDASO, a heterobifunctional sulfoxide-based (yellow box) cleavable photo-cross-linker. Shown on the right are the heterobifunctional NHS-carbamate photo-cross-linkers, DizSEC and DizSPC, reported in this study. (B) Reaction of the NHS-carbamate functional group in DizSEC or DizSPC with lysine or N-terminal amines generates ureas. Following UV irradiation, the diazirine forms a carbene, which rapidly reacts with proximal proteins to form covalent cross-links. (C) Ureas are doubly fissile under HCD (across each N-carbonyl bond) and generate two distinct sets of fragment ions (green and purple dashed boxes), generating amine and isocyanate fragments on either peptide product (nomenclature used for these fragment ions shown). These cross-linked peptide fragments form characteristic doublets in the MS spectrum. (D) Representative XSM generated by the treatment of BSA with DizSEC, along with a structure (PDB: 4F5S) illustrating identified cross-links. The spectrum contains all four possible products from cleavage of the MS-cleavable cross-link (that is, the modified parental peptides).

collision-induced dissociation or higher-energy collision-induced dissociation (HCD). These cross-linkers feature a functional group (cyan and yellow boxes in Figure 1) which is readily broken, splitting the cross-linked peptide into two modified linear peptides. Since the functional group is also symmetrical, fissure can occur in one of two ways (green and purple dotted lines), which further facilitates the identification of spectra of cross-linked peptides by generating doublets with a characteristic separation (Figure 1C). These features collectively enable MS-cleavable cross-linkers to mitigate the N^2 problem by reducing the database size against which spectra are searched and allowing complex samples to be interrogated by XL-MS. While numerous functional groups have been reported to be MS-cleavable,[14] sulfoxides and ureas are among the most frequently employed.[15,16] With these chemical tools, several proteome-wide cross-linking studies have been reported, such as on *Bacillus subtilis* extracts,[17] *Escherichia coli* extracts,[17,18] yeast extracts,[19] *Drosophila* embryonic extracts,[16] *Mycoplasma pneumoniae* cells,[5,20] murine mitochondria,[21] murine heart tissues,[22] human adenocarcinoma extracts,[23] and in HEK-293 cells.[24,25]

Photo-cross-linkers [also called photo-affinity labels (PALs)] possess chemically inert photo-reactive functional groups, which upon UV irradiation, generate highly reactive moieties that rapidly and covalently capture nearby proteins.[26,27] Though azides, benzophenones and diazirines are the most utilized PALs in biological systems, recent studies have reported aryl-tetrazoles and *o*-nitrobenzyl alcohols as PALs to capture PPIs.[28,29] Diazirines are "zero-length" cross-linkers that are stable in aqueous media at room temperature, are synthetically tractable in a broad range of architectures, and offer predictable photo-cross-linking products; hence, they have emerged as a popular "universal" photo-affinity label.[30−32] Additionally, diazirines are photo-activated at relatively benign UV wavelengths (~365 nm), avoiding protein damage that can occur at shorter wavelengths (<300 nm).[33] Upon photo-activation, diazirines expel N_2 and form reactive carbenes that can insert across a diverse array of chemical bonds, including N−H, C−H, O−H, and S−H. Aliphatic carbenes evolved from

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

diazirines have short lifetimes (ns−μs) and are reactive toward any amino acid in proximity. Hence, this combination of properties enables diazirine-based photo-cross-linkers to capture rapid or transient interactions and significantly expands "cross-linkable" chemical space, given that most chemical cross-linkers attach to a restricted profile of residue-types (primarily Lys or Cys).[27]

Heterobifunctional photo-cross-linkers such as succinimidyl-4,4′-azipentanoate (SDA) (Figure 1A) contain an amine-reactive NHS-ester and a diazirine (green box) separated by an alkyl chain. This family of cross-linkers was initially introduced to assist investigations of PPIs by covalently coupling interacting proteins to a target protein of interest.[34−37] They utilize a "plant-and-cast" strategy whereby the amine-reactive NHS-ester[38] plants the compound onto a target protein, casting the photo-cross-linking moiety to react with proximal residues upon subsequent UV irradiation. The cross-links generated by heterobifunctional photo-cross-linkers can also be sequenced with mass spectrometry; however, identifying photo-cross-linked peptides is technically challenging because (i) their higher reactivity results frequently in quenching and consequently lower cross-linking yields; and (ii) they can react with all residue-types through a variety of linkages.[39,40] Hence, photo-cross-linkers have found some applications in structural studies to generate interaction maps spanning diverse residue types (yielding orthogonal and complementary information to what is obtained with conventional chemical cross-linkers);[41,42] however, the challenges associated with photo-cross-linking have resulted in limited applicability to proteome-wide experiments. One study by Iacobucci and co-workers demonstrated that diazirines can generate MS-cleavable ester moieties when their photoproducts cross-link with Asp or Glu residues, facilitating identification on several purified proteins (e.g., calmodulin and GCAP-2).[43] Moreover, a recent study by Gutierrez et al. reported a series of diazirine-based cross-linkers bearing the MS-cleavable sulfoxide moiety (SDASO, as shown in Figure 1A).[41] These cross-linkers have advanced the study on single purified proteins (or protein complexes), however are not without limitations. Ester linkages are MS-cleavable but do not leave behind modifications upon cleavage. SDASO's MS-cleavable cross-link does produce mass modifications after cleavage but is singly fissile[44] and hence does not generate signature doublets (e.g., Figure 1C). We conjectured that these spectral features could be important for facilitating accurate identification of cross-linked spectra from complex proteome-wide samples.

To the best of our knowledge, a photo-cross-linker has not yet been deployed to identify cross-links in a proteome-wide experiment. We reasoned that a heterobifunctional photo-cross-linker that is MS-cleavable in a doubly fissile manner (thereby producing doublet peaks) could improve confidence in reliably identifying cross-links from highly complex samples. We were inspired by the compound NNP-9 (Figure 1A), a trifunctional chemical cross-linker that features an NHS-carbamate group (red box).[45] Following this concept, we designed two heterobifunctional cross-linkers like SDA, in which the NHS-ester moiety is replaced with an NHS-carbamate. Following a nucleophilic attack by lysine, NHS-carbamates yield urea functional groups at the initial attachment site (Figure 1B). This group would possess doubly fissile MS-cleavability, which could facilitate spectral search (Figure 1C), though this latent capacity of NHS-carbamates has not been previously exploited.

Here, we report the synthesis and evaluation of DizSEC and DizSPC, photo-cross-linkers featuring NHS-carbamates and aliphatic diazirines with ethyl and propyl spacer arms, respectively, separating the chemical and photo cross-linking moieties. In the following, we demonstrate that DizSEC and DizSPC can generate readily identifiable cross-links in pilot studies on purified proteins. We further demonstrate that DizSEC and DizSPC are cell-permeable and develop a method to perform rapid *in vivo* photo-cross-linking on whole *E. coli* cells. Next, we show that these cross-linkers' doubly fissile MS-cleavability, in combination with an experimental and computational workflow developed here, enables confident identification of cross-links in the *E. coli* proteome. The data provide a partial interaction map of *E. coli* with residue-level resolution, capturing some previously unknown PPIs and providing some structural insights that are discussed (along with potential biological implications) in the following. Therefore, the compounds and methods described in this work provide a foundation to perform proteome-wide photo-XL-MS, which we hope will become a valuable tool to elucidate the molecular sociology of the cell with high temporal, spatial, and compositional resolution.[46,47]

EXPERIMENTAL METHODS

Synthetic Methods. Heterobifunctional photo-cross-linkers DizSEC and DizSPC were synthesized in a seven-step protocol from commercial reagents with yields of 86.7% (111 mg) and 85.9% (116 mg), respectively. Detailed synthetic procedures and characterization are reported in Supporting Information (Scheme S1, Figures S1−S7).

Cross-Linking Lysozome and BSA with DizSEC and DizSPC. Lysozyme and bovine serum albumin (BSA) were purchased from Sigma-Aldrich, were prepared in reaction buffer (50 mM Na-HEPES pH 8.0, 100 mM NaCl, 10 μM final concentration protein) at a 50 μL scale to which 0.5 μL of freshly prepared photo-cross-linkers DizSEC, DizSPC, or sulfo-SDA (as positive control) was added as stocks in DMSO (100 mM or 1 M stock concentration) to a final concentration of 1 mM or 10 mM. Some samples received 4-dimethylaminopyridine (DMAP) to a final concentration of 0.1 or 1 mM.[48,49] Samples were gently mixed by inversion, then incubated at room temperature on an end-over-end rotator for 2 h (10 rpm). Chemical cross-linking reactions were quenched by the addition of Tris−HCl pH 8.0 (100 mM, final concentration) followed by incubation for 15 min at room temperature. Samples were transferred to a 1.0 mm path length quartz cuvette and photolyzed for 5.0 s with a 4.0 cm^2 Hönle LED Spot W UV (365 nm)-LED array rated with a 14 W/cm^2 max output (20 or 80% intensity). The distance between the cuvette and the LED head was minimized (ca. 1−2 mm) by affixing the cuvette with the tape to the array. Tables S1 and S2 (Supporting Information) catalogue various experimental cross-linking conditions used for the lysozyme and BSA samples.

Proteome-Wide Cross-Linking of *E. Coli* Extracts. *E. coli* cells (K12 subst. MG1655) were grown overnight in LB (Luria-Bertani) media (5 mL) at 37 °C, 220 rpm. Overnight cultures were diluted into 50 mL of LB media with an initial OD_{600} of 0.05 until a final OD_{600} of 1.0 was reached. The cultures were pelleted in a 5910R swing-bucket benchtop centrifuge (Eppendorf) at 4000*g* for 10 min (4 °C), and the supernatant was discarded. The cell pellet was resuspended in 1 mL of HEPES reaction buffer and then lysed by sonication

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

on water−ice with an FB120 probe sonicator (Fisher Scientific) at 45% amplitude for 10 min in 10 s/10 s on−off cycles. *E. coli* lysates were centrifuged at 14 000*g* for 10 min on a 5425R benchtop centrifuge (Eppendorf) to clarify the samples. BCA assay (Pierce) was performed following a manufacturer's protocol to determine the protein concentration. Samples were normalized to 1.0 mg/mL protein with reaction buffer. The cross-linking reaction was performed on a 1 mL scale of the normalized extracts. DizSEC or DizSPC (10 mM final concentration) were added from 1 M stocks in DMSO. Some samples were also supplied with 1 mM DMAP (final concentration) as a cross-linking catalyst. Samples were gently mixed by inversion, then incubated at room temperature on an end-over-end rotator for 2 h (10 rpm). Chemical cross-linking reactions were quenched by the addition of Tris pH 8.0 (100 mM, final concentration) followed by incubation for 15 min at room temperature. Samples were transferred to a 1.0 mm path length quartz cuvette (Starna Cells, PN: 21-Q-1) and photolyzed for 5.0 s with a 4.0 cm^2 Hönle LED Spot W UV (365 nm)-LED array rated with a 14 W/cm^2 max output (20 or 80% intensity). The UV-LED array was connected to a 200 W LED PowerDrive (Hönle).

In-Cell Photo-Cross-Linking of *E. coli*. Overnight and day cultures were prepared identically to that described above. 50 mL of LB cultures (OD_{600} 1.0) was pelleted in a 5910R swing-bucket benchtop centrifuge (Eppendorf) at 4000*g* for 10 min (4 °C), and the supernatant was discarded. Cells were resuspended and washed in 20 mL of HEPES reaction buffer four times. After the final centrifugation, cells were resuspended in 1 mL of HEPES reaction buffer, and the NHS-carbamate coupling reactions were performed by the addition of DizSEC or DizSPC (10 mM final concentration). Cell suspensions were gently mixed by inversion, then incubated at room temperature on an end-over-end rotator for 2 h (10 rpm). The cross-linking reaction was quenched by the addition of Tris pH 8 (100 mM, final concentration) and incubation for 15 min at room temperature. Cell suspensions were diluted 10-fold by the addition of 9 mL of HEPES reaction buffer before photolysis to prevent excessive scattering.

The photolysis apparatus consisted of the LED Spot W UV-LED array with a 1 mm pathlength quartz flow-cell (Starna Cells, PN: 45-Q-1) affixed directly to the LED array surface with tape. The flow cell was connected to a Masterflex L/S model 77200-50 peristaltic pump via 1.6 mm ID C-Flex biocompatible tubing (Masterflex). The tubing and flow cell were primed with water at the max speed to remove bubbles. The cell suspension was pumped through the flow cell at 50 rpm (approximately 1 s exposure time), collected, and centrifuged at 4000*g* for 10 min. The supernatant was discarded, and cross-linked cells were resuspended in 1 mL of HEPES reaction buffer, then lysed by sonication, and normalized to 1.0 mg/mL protein using the BCA assay as described above.

MS Sample Preparation. Cross-linked protein mixtures (typically 1 mL of 1.0 mg/mL) were denatured by addition of urea (1.92 g) and ammonium bicarbonate (Ambic, 15.8 mg) and diluted to a final volume of 4 mL by addition of Millipore water to generate samples that are 0.25 mg/mL protein, 8 M urea, and 50 mM Ambic (pH 8). These samples were reduced by the addition of dithiothreitol (DTT, Sigma-Aldrich) to 10 mM final concentration and incubation at 37 °C with agitation (700 rpm) in a benchtop ThermoMixer (Eppendorf) for 30 min. Samples were then alkylated by the addition of iodoacetamide (IAA, Acros Organics) to 40 mM final concentration and incubation at room temperature in the dark for 45 min. Samples were either (i) diluted with 3 volumes of 50 mM Ambic and digested with trypsin (1:50 enzyme:protein w/w ratio, Pierce) overnight (ca. 16 h) at 25 °C, 700 rpm in a ThermoMixer or (ii) digested serially—first with Lys-C (1:100 enzyme:protein w/w ratio, Pierce) for 2 h at 37 °C, then diluted with 3 volumes of 50 mM Ambic followed by second digestion with trypsin overnight (ca. 16 h). Desalting of protein digests was performed using Sep-Pak C18 cartridges (Waters).

Lysozyme and BSA samples were desalted with C18 cartridges (Waters Corporation). First, digests were acidified with HPLC-grade trifluoroacetic acid (TFA, Alfa Aesar) to a final concentration of 1% v/v and diluted to 1 mL with 0.5% TFA in Optima liquid chromatography/mass spectrometry (LC/MS) grade water (Fisher chemical, Buffer A). Cartridges were placed onto a vacuum manifold, conditioned with 2 × 1 mL of 0.5% TFA in 80% Optima LC/MS grade acetonitrile (Fisher chemical, Buffer B), and equilibrated with 4 × 1 mL of Buffer A. Acidified digests were loaded onto the cartridges under a reduced vacuum (such that it took approximately 5 min for 1 mL to passage through the sorbent) and then washed with 4 × 1 mL of Buffer A. The cartridges were placed on top of 15 mL conical tubes, and 1 mL of Buffer B was pipetted onto the resin. The cartridges were gently eluted under centrifugal force in swing-buckets spun at 350 rpm in a 5910R centrifuge (Eppendorf) for 3 min and then reduced to dryness in a Vacufuge centrifugal concentrator (Eppendorf).

E. coli extracts (either cross-linked in lysate or *in cellulo*) were desalted analogously, except using high-capacity C18 cartridges (Waters Corporation). Samples were acidified as described above. High-capacity cartridges were placed onto a vacuum manifold, conditioned with 2 × 10 mL of Buffer B, and equilibrated with 4 × 10 mL of Buffer A. Acidified digests were loaded onto the cartridges under a reduced vacuum (such that it took approximately 5 min for the ca. 12 mL to passage through the sorbent) and then washed with 4 × 10 mL of Buffer A. The cartridges were placed on top of 50 mL conical tubes, and 5 mL of Buffer B was pipetted onto the resin. The cartridges were gently eluted under centrifugal force in swing-buckets spun at 350 rpm on a 5910R centrifuge for 5 min and then reduced to dryness using a Vacufuge centrifugal concentrator. Dried peptides were stored at −80 °C until further use.

Enrichment and Fractionation of Cross-Links by Size Exclusion Chromatography. Cross-linked peptides derived from *E. coli* extracts and in-cell cross-linked samples were enriched by peptide size exclusion chromatography (SEC). Dried peptides were reconstituted in 250 μL of 3% acetonitrile and 0.1% TFA with vigorous sonication. Peptides were fractionated on an AKTA Go FPLC (Cytiva) with a Superdex 30 10/300 GL gel filtration column. Chromatography used an isocratic method with solvent A (3% acetonitrile, 0.1% TFA) and a flow rate of 0.4 mL/min (approximate column pressure 2.5 MPa) with a column temperature of 4 °C. Prior to fractionation, the column was equilibrated with solvent A for 2 h. 250 μL peptides was injected onto the column after being loaded into a 500 μL loop, and 0.2 column volumes (the void volume) were passed to waste without being collected. Elution volumes were set to 0.5 mL for fractions 1−12, with collection beginning 8 mL after injection; elution volumes were set to 1.0

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

mL for fractions 12−24. Out of 24 fractions, 4−20 were found to be enriched with cross-linked peptides, and these fractions were selected for LC−MS/MS analysis. A_{280} traces from the SEC enrichment or proteome-wide cross-linking experiments can be found in Figure S12. Fractions were reduced to dryness in a Vacufuge centrifugal concentrator and stored at −80 °C until further use.

LC−MS/MS. This study utilized an UltiMate3000 UHPLC system (Thermo Fisher) coupled with a Q-Exactive HF-X Orbitrap mass spectrometer (Thermo Fisher). Peptide fractions were resuspended in 30 μL of 0.1% Optima formic acid (FA) in Optima water, vigorously vortexed, and sonicated for 5 min. The amount of peptide material in each resuspended sample was approximately quantified with a NanoDrop One^C microvolume UV−vis spectrophotometer (Thermo Fisher Scientific), and typically ∼1 μg of the peptide was injected onto the LC−MS system for each run. If a fraction contained insufficient material to inject 1 μg of peptide, 10 μL of the sample (one-third) was injected instead. The column and trap cartridge temperatures were set to 40 °C, and the flow rate was set to 300 μL/min. Mobile phase solvent A comprised 0.1% FA, 2% Optima acetonitrile (Fisher Scientific), and 98% Optima water. Mobile phase solvent B comprised 0.1% FA in 98% Optima acetonitrile and 2% Optima water (Fisher Scientific).

The injected sample was loaded onto a trap column cartridge (Acclaim PepMap 100, C18, 75 μm × 2 cm, 3 μm, 100 Å column) and washed with solvent A for 10 min. Next, the trap column was switched to be in line with the separating column (Acclaim Pepmap RSLC, C18, 75 μm × 25 cm, 2 μm, 100 Å column). The sample was separated in a linear gradient from 2% B to 35% B over 100 min, to 40% B over 25 min, and to 90% B over 5 min. The end of the LC method employed a saw-tooth gradient to remove the sample residue from the resolving column.

XL-MS data were acquired in the positive-ion mode with each full MS scan using a mass range of 350−1500 m/z. The resolution of the full MS scans was set to 120k, the automatic gain control (AGC) target was set to 3×10^6, the maximum injection time (IT) was set to 100 ms, and a stepped HCD fragmentation scheme was analyzed using a three-tiered stepped normalized collisional energy scheme. Various NCEs were tested to optimize these parameters for the different experiment types and are documented in Table S1; however, 22, 25, and 28% proved acceptable for most of the experiments. After each full MS scan, 10 data-dependent MS^2 scans were collected at a resolution of 15k, an AGC target of 2×10^5, a minimum AGC target of 8×10^3, a maximum IT of 250 ms, an isolation window of 1.0 m/z (some lysozyme samples used an isolation window of 1.4 and 2.0 m/z; see Table S1), and a dynamic exclusion time of 60.0 s. Ions with charges less than +3 or greater than +8 and isotopomeric peaks were excluded from data-dependent MS^2 scans.

MeroX Data Analysis Workflow. Raw files were converted to the mzML format as centroid spectra using the MSConvert GUI from the ProteoWizard toolkit,[50] and a spectral search for cross-linked peptides was conducted with MeroX[51] (version 2.0.1.4). For both DizSEC and DizSPC, protease cleavage sites were allowed after Arg and Lys, with Lys cleavage blocked by Pro, and a maximum of three missed cleavages were allowed. The minimum peptide length was set to three residues, and the maximum peptide length was set to 30 residues. Static modifications included the carbamidomethylation of cysteine (C to B), and variable modifications included a maximum of 1 oxidation of methionine (M to m) per peptide (greater than 1 prohibitively expands the search space). The precursor precision was set to 10.0 ppm, and the fragment ion precision was set to 20.0 ppm. Mass recalibration for each raw file was determined manually by performing a separate search on each raw file for linear peptides in Proteome Discoverer 2.4 (Thermo Fisher Scientific) without spectrum recalibration, and the median $\Delta m/z$ (typically between 5−8 m/z) was relayed to the MeroX settings file for each individual search. The lower mass limit was set to 350 m/z, the upper mass limit was set to 1500 m/z, the signal-to-noise ratio was set to 2.0, b- and y-ions only were enabled, the minimum number of fragments per peptide was set to 3, the minimum MS^1 charge was set to 3, and the deisotoping mode was selected. We used a pre-score intensity of 10% and an FDR cutoff of 1%. Decoys were generated by shuffling sequences but keeping protease sites within the supplied FASTA files. Typical scores of identified peptides ranged between 20 and 120. Lysozyme and BSA samples used quadratic mode, while cross-links derived from *E. coli* (both from cross-linking in lysate or *in cellulo*) utilized the proteome-wide mode developed by Götze and co-workers.[16] For the lysozyme and BSA samples, a FASTA file was prepared that contained the sequence of hen egg white lysozyme (P00698) or BSA (P02769). For the *E. coli* lysates and in-cell cross-linked samples, the reference FASTA for *E. coli* K-12 substrain MG1655 (UP000000625) was used.

To encode the cross-linkers in MeroX, for the DizSEC-cross-linked samples, the composition mass was set to C_5NOH_7 (97.05276383 Da). The first specificity site for the NHS-carbamate was Lys and the N-terminus, and the second specificity site for the diazirine was set to all amino acids, both termini, carbamidomethylated cysteine, and oxidized methionine. Modifications at site 1 had composition masses of COH_{-2} (25.97926456 Da; lys_{CO}) and the unmodified peptide (i.e., 0; lys_{pep}). Modifications at site 2 had composition masses of C_4NH_9 (71.07349927 Da; non_{amine}), C_5NOH_7 (97.05276383 Da; $non_{isocyanate}$), and unmodified peptide (0; discussed more in the following).

For the DizSPC-cross-linked samples, the composition mass was set to C_6NOH_9 (111.06841389 Da). Specificity sites and modifications at site 1 were the same as in the DizSEC settings. Modifications at site 2 had composition masses of C_5NH_{11} (85.08914933 Da; non_{amine}), C_6NOH_9 (111.06841389 Da; $non_{isocyanate}$), and unmodified peptide (0).

For double-validation "entrapment" searches, a FASTA file was prepared which contained the 1500 most abundant *E. coli* proteins [as based on pooling fractions, performing a standard linear peptide search, and sorting based on peptide-spectrum matches (PSMs)] as well as 1500 randomly chosen yeast proteins (taken from UP000002311). Search and FDR settings were otherwise identical to the production searches.

Method files for running these searches in MeroX and an executable script to automate job submission are provided (10.5281/zenodo.7778901).

XiSearch and XiFDR Data Analysis Workflow. Raw files were converted to the .mgf file format with mass calibration conducted by MSConvert (run via a script, also available on the Zenodo deposition). MSConvert was directed to identify the optimal mass shift by comparing spectra against a FASTA file containing the 400 most abundant proteins in *E. coli* (as based on a ranked list of number of PSMs from a separate search of linear peptides conducted in Proteome Discoverer).

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

Recalibrated .mgf files, settings files, and FASTA files were moved to the Rockfish computing cluster at Advanced Research Computing at Hopkins (ARCH) via the file transfer program, WinSCP (Version 5.21.7). Individual recalibrated .mgf files were searched with XiSearch (Version 1.7.6.7) for cross-linked peptides using two search settings (both inclusive and conservative), as described in the text and shown in Table S11 (Supporting Information). All versions of the settings included the variable modifications of the oxidation of methionine (+15.99491463 Da) and the cross-linker loop-link modification at Lys and the N-terminus (+97.05276383 Da for DizSEC or +111.06841389 for DizSPC) as well as the fixed modification of the carbamidomethylation of Cys (+57.021464). XiSearch was set to search only b- and y-ions, and the digestion enzyme was set to "constrained" (blocked by proline) trypsin. Water loss (−18.01056027 Da) was allowed at Ser, Thr, Asp, Glu, and the C-terminus. Ammonia loss was allowed at Arg, Lys, Gln, Asn, and the N-terminus. DizSEC was encoded as a cross-linker with a mass of 97.05276383 Da with the first linked amino acid as Lys or the N-terminus and all canonical amino acids and the termini allowed as the second linked amino acid. DizSEC cleavage products were encoded as modifications of 25.97926456 Da at Lys or the N-terminus (lys_{CO}), 0 Da at Lys or the N-terminus (lys_{pep}), and 71.07349927 Da (non_{amine}) or 97.05276383 Da ($non_{isocyanate}$) at any amino acid. DizSPC cleavage products were encoded as losses of 25.97926456 Da at Lys or the N-terminus (lys_{CO}), 0 Da at Lys or the N-terminus (lys_{pep}), and 85.08914933 Da (non_{amine}) or 111.0684189 Da ($non_{isocyanate}$) at any amino acid.

For the "inclusive" search settings, three missed cleavages and three modifications per peptide were allowed. The inclusive settings used wider tolerances of 10.0 ppm at the MS^1 level and 20.0 ppm at the MS^2 level. The inclusive search used the full *E. coli* FASTA file, consisting of 4448 proteins. These searches were computationally expensive and required high-performance computational clusters with high memory allocations (500 GB) to run to completion.

The conservative settings allowed only one missed cleavage and one modification per peptide. These settings used a tighter mass tolerance of 5.0 ppm at the MS^1 level and 15.0 ppm at the MS^2 level. The conservative search used a smaller FASTA file that contained the top 2000 proteins by number of PSMs as identified by Proteome Discoverer. These searches could be run on a conventional laboratory workstation (ours has 8 cores with 2.934 GHz processors and 64 GB RAM).

Searches were run on Rockfish using a multi-submission script (provided on Zenodo deposition) using 500 GB of RAM, which was critical for completion of searches using the inclusive settings. Search output files that corresponded to a single experimental condition (e.g., all fractions that were collected using the DizSEC cross-linker for *in cellulo* cross-linking) were concatenated using a script (available on Zenodo). The concatenated data file was provided to the XiFDR GUI along with the appropriate settings file and FASTA. Several validation settings were tested (see Supporting Information, Table S12); the setting we settled on was XiFDR's default with two modifications: a prefilter for doublet counts (CCfragmentdoubletcount) to be greater than 0 and boosting on "between" (interprotein) cross-links. Figure S13 shows a flow−chart diagram illustrating the computational workflow.

Network Analysis. MeroX data and XiFDR data were exported from their respective programs in the XiView format, including identification and sequence files. These files were uploaded to XiView[52] with a Met oxidation modification mass of 15.99491 Da and a Cys carbamidomethylation modification mass of 57.02146 Da. Each set of the experimental data was aggregated together on the XiView server to form network diagrams. Protein Data Bank (PDB) codes were imported into XiView, and Euclidian cross-link distance measurements were exported into PyMol scripts. Structures with mapped cross-links were generated in PyMOL by importing the PDB file and running the appropriate PyMOL command script.

Protein Docking. HADDOCK .tbl restraint files were exported from XiView for each complex to be docked. Alphafold or PDB files for the appropriated structures were edited with PDBTools to remove unwanted heteroatoms and headers, then submitted to the HADDOCK 2.4 webserver.[53,54] Only the polypeptide chains involved in the cross-linking interaction were selected for docking. Cross-links were provided as ambiguous restraints with an upper distance bound of 30 Å, except for the docking of ManZ and ManY, for which the two cross-links were provided as unambiguous restraints. All other settings used for docking were the default HADDOCK settings. Models for docked structures were either taken from the Protein Data Bank if there was an available *E. coli* structure (the highest resolution was selected if there were multiple) or from AlphaFold if there was not. No violations were observed in any of the docked structures. Docked structures were then aligned with their original (full) PDB or AlphaFold structures in Pymol.

RESULTS

Piloting Photo-Cross-Linkers on Purified Proteins. Initially, we combined DizSEC with N-Boc-L-lysine in a HEPES buffer (pH 8.0) to measure the kinetics of the acyl transfer reaction (Figure S8). This confirmed that the NHS-carbamate was a comparable (though somewhat slower) electrophile to the more standard NHS-ester, with the reaction coming to completion at 60 min. Next, we screened several different experimental conditions with DizSEC and DizSPC on a single protein (lysozyme). Parameters screened included concentration of the cross-linker, power of the UV-LED array, presence or absence of a Lys-C pre-digestion step, the addition of DMAP as an acyl-transfer catalyst, and variable collisional energies. The results of these experiments are compiled in Tables S1 and S2, and the cross-links identified are mapped in Figure S9. Preliminary results confirmed the success of the cross-linking experiment, and cross-linked peptide identification was not very sensitive to the power of the UV lamp, the presence or absence of Lys-C or DMAP, or cross-linker concentration—suggesting that these compounds function as robust protein cross-linkers. As expected, the identification of cross-linked peptides was explicitly dependent on treatment with a chemical cross-linker (Tables S1, S2).

To test that one of these conditions would generalize well to an unrelated protein, we repeated this experiment on BSA (Figure S10). Figure 1D shows an example of one of the assigned spectra from MeroX on one cross-linked peptide from BSA. Altogether, we identified 249 spectra that matched to intraprotein cross-linked peptides (and 1689 intrapeptidal cross-links), corresponding to 143 unique intraprotein cross-links. The cross-link-spectrum match (XSM), as shown in Figure 1D, corresponds to a lysine (K476) cross-linked to a leucine residue (L456) with support from 73 matched fragment ions. In this case, all four of the characteristic

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

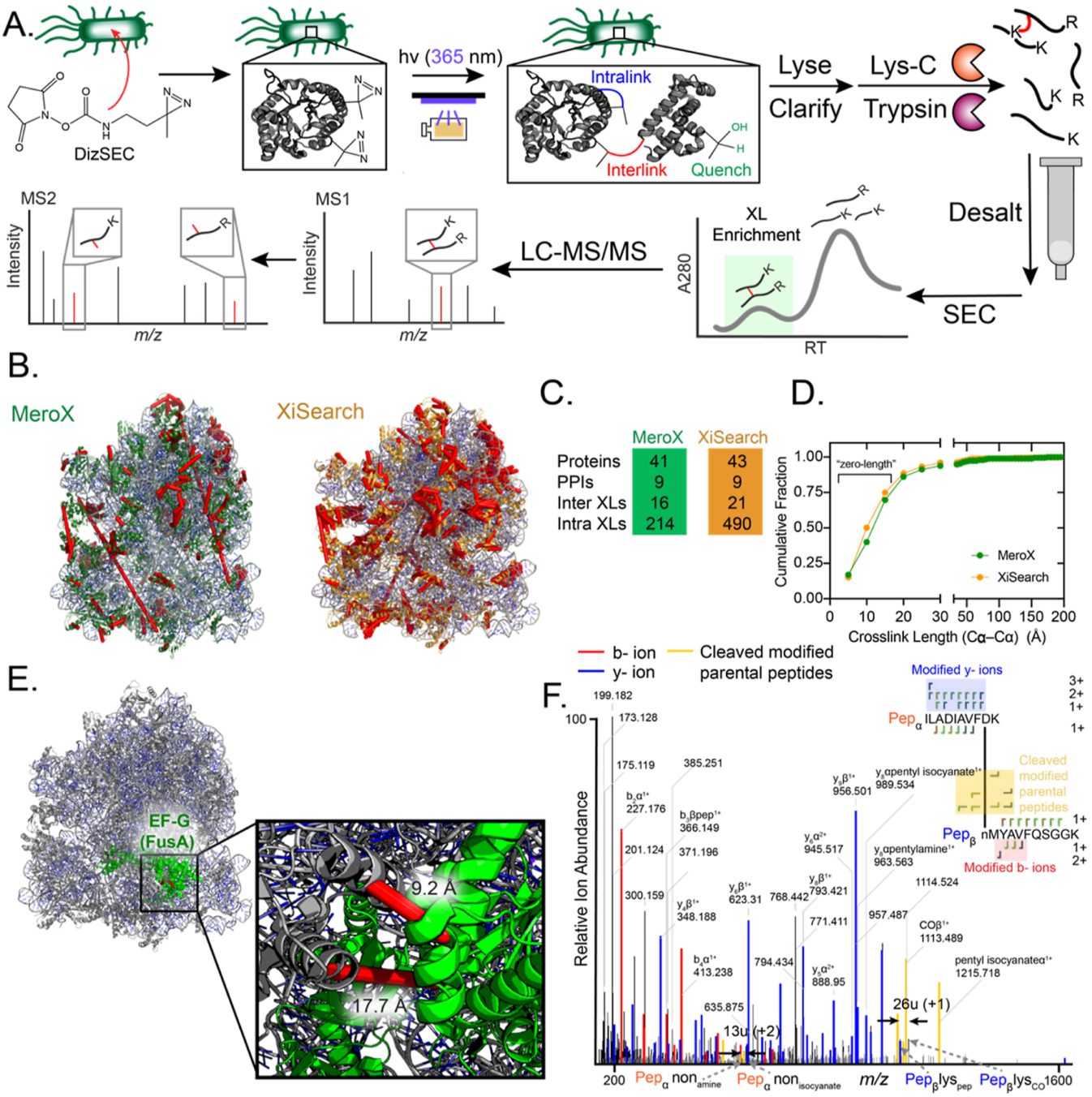


Figure 2. Proteome-wide identification of photo-cross-links. (A) Experimental workflow to prepare cross-links for LC−MS/MS identification using DizSEC (or DizSPC) on *E. coli* extracts. (B) Identified cross-links from MeroX and XiSearch mapped onto the 70S *E. coli* ribosome (PDB: 7CPJ). (C) Summary data showing the number of cross-links mapped onto the ribosome for both search algorithms. (D) Cumulative distribution plot showing the Cα−Cα distances of the cross-links identified. The majority of cross-links identified in both search algorithms agree with a cryo-EM structure of the ribosome (<15 Å). (E) Structural representation of cross-links mapped onto the 70S ribosome complexed with EF-G (PDB: 7N2C). The inset shows a detailed view of these cross-links, which have lengths that agree with the structure. (F) Representative interprotein XSM of RplT (αPep) cross-linked with RplU (βPep) with (DizSPC).

fragment ions in which fissure occurred at the cross-link (but nowhere else, releasing modified linear peptides) are detected, resulting in two doublets each separated by 13 *m/z* (for the 2+ charge state). The fragmentation at the cross-link leaves behind characteristic modifications on the non-lysine cross-linking site (an alkyl amine and alkyl isocyanate, cf. Pep_α on Figure 1C). These are frequently found in the b- and y-ion series and facilitate the identification of the attachment site selected by the photo-activated diazirine. As shown in Figure 1D, y-ions of the α-peptide containing the isocyanate modifications are apparent from y_{13} down to y_7, but starting with y_6 the isocyanate modification disappears. Hence, the attachment site can be uniquely assigned to L8. This finding confirms that the urea generated at the DizSEC cross-link junction (formed by acyl transfer from lysine to the NHS-carbamate) can fragment more readily than standard peptide bonds, similar to the popular cross-linker DSBU.

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

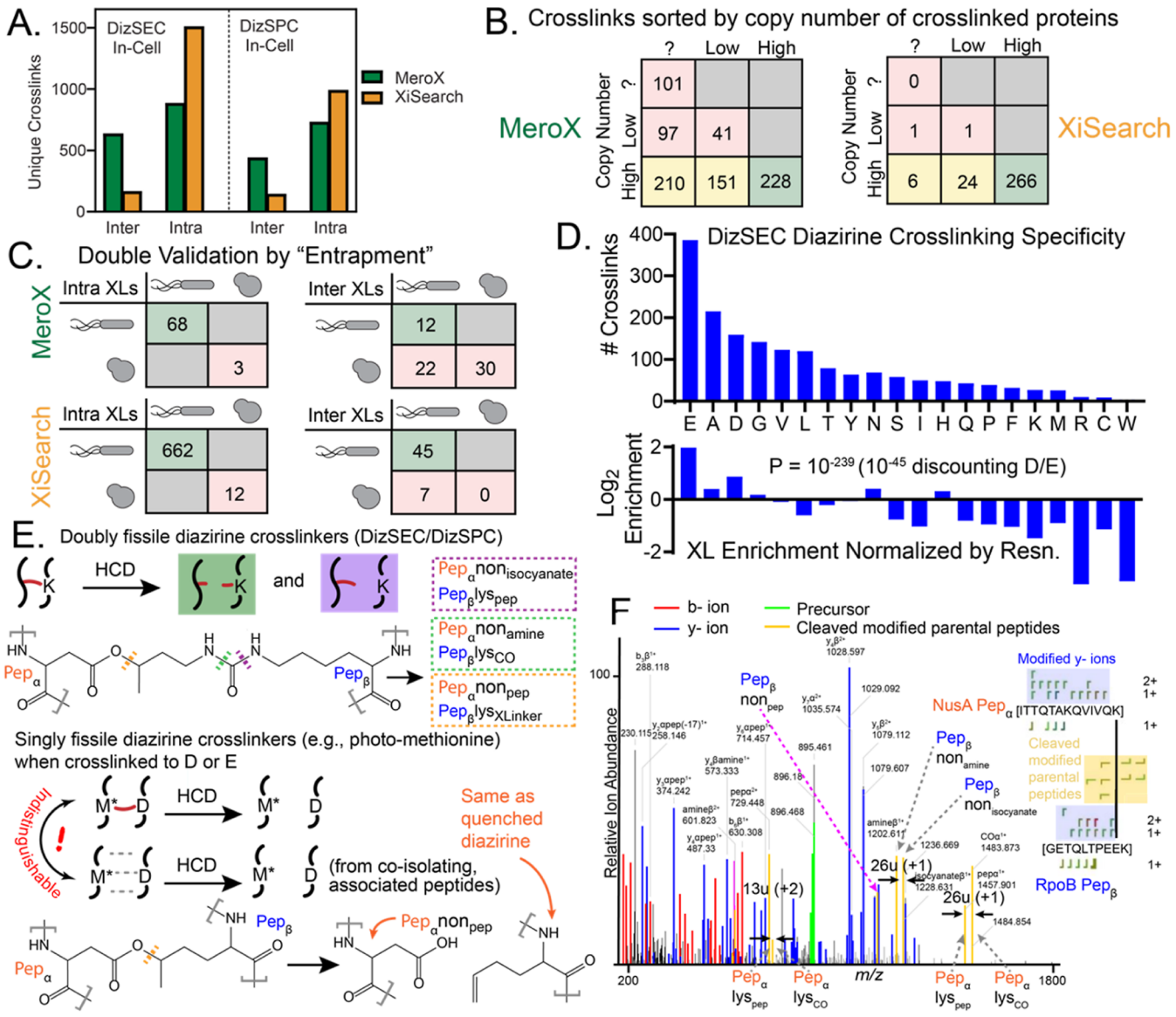


Figure 3. Assessment of false discovery across algorithms. (A) Summary data comparing the number of unique cross-links identified by MeroX and Xi in cross-linking *in cellulo* experiments with DizSEC/DizSPC. (B) MeroX finds many interprotein cross-links between proteins with low or unknown copy number; this is not recapitulated in the Xi workflow. (C) MeroX often finds impossible cross-links when a double validation is performed. Erroneous yeast cross-links make up 22% of identifications for this fraction and 81% of interprotein identifications. Conversely, there are far fewer erroneous XiSearch IDs for the DizSEC in-cell fractions. (D) Number of cross-links identified in the Xi workflow for the DizSEC *in cellulo* experiment, sorted by the identity of the "non-lysine" attachment site. Below, the enrichment for each residue normalized to its frequency in the proteome. (E) Doubly fissile cross-linkers leave a modification on three of the four possible cleavage products, unlike scenarios in which simple diazirines (e.g., photo-Met) attach to Asp or Glu, which do not leave modifications on the fragments ions. These unmodified cleavage products are indistinguishable from associated co-isolating peptides. (F) MS2 spectrum showing a NusA peptide (at K117) cross-linked with an RpoB peptide (at E899) via DizSEC. Three sets of modified, cleaved, parental peptide doublets are highlighted in yellow, resulting from cleavage at the urea moiety. The pink arrow points to a product formed upon cleavage at the ester, which produces the unmodified nonlysine peptide.

The number of cross-links identified with DizSEC on lysozyme was similar to that of the commercially available non-cleavable photo-cross-linker, sulfo-SDA. Hence, for studies on a single protein, MS-cleavability is not a necessity (Tables S1, S2). However, with DizSEC, we typically identify more fragment ions containing the cross-link site, so that it is easier to accurately pinpoint the non-lysine attachment site with DizSEC compared to sulfo-SDA (see Figure S11, showing the same peptide-pair as in Figure 1D, but cross-linked with sulfo-SDA). This is consistent with the recent finding that MS-cleavable cross-linkers improve backbone fragmentation.[55] DizSPC had notably fewer cross-links on lysozyme but gave comparable results to DizSEC on BSA (Table S2). Cα−Cα cross-link distances for lysozyme and BSA (based on X-ray structures with PDB: 3LYZ and 4F5S) are provided in Tables S3 and S4, respectively. The median Cα−Cα cross-link distances were very short for both model proteins (17.3 and 13.0 Å for lysozyme and BSA, respectively), confirming the capacity of the diazirine to function as a zero-length cross-linker, and suggesting that our assignments of the identity of the cross-linked residue on the non-lysine-containing peptides are generally accurate.

Developing a Workflow for Proteome-Wide Photo-Cross-Linking *In Cellulo*. To investigate if DizSEC and

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

DizSPC were capable of probing protein structure and interaction networks in their native cellular context, we incubated *E. coli* with the compounds after washing away the primary amine-rich LB media and performed photolysis on live cells. These experiments take advantage of the high-power UV LED array, coupled to a flow cell (Figure 2A, see also Experimental Methods) such that each cell is illuminated for approximately 1 s. With such rapid irradiation, we anticipated that the experiment would capture a snapshot of the cell's "state" during photolysis and trap transient interactions. This photolysis method reduces diazirine cross-linking time substantially from typical irradiation methods (several minutes) and hence reduces the likelihood that results would be influenced by how cells might behave downstream to UV stress. Following rapid photolysis, cells were collected and lysed. The extracts were subject to double in-solution digest (first with Lys-C in 8 M urea, then with trypsin in 2 M urea) and the peptides cleaned up with solid phase extraction. Anticipating high sample complexity, we enriched for cross-linked peptides and fractionated them by size-exclusion chromatography. 16 fractions were collected that were deemed likely to contain cross-links (Figure S12) and were subsequently analyzed by LC−MS/MS, as described in the Experimental section.

Because the complexity of these samples is greater than most cross-linking experiments (any lysine in any *E. coli* protein can cross-link to any other *residue* in the *E. coli* proteome), we wanted to test and compare search and validation workflows with two leading software platforms for XL-MS, MeroX (developed by Sinz and colleagues),[56] and Xi (developed by Rappsilber and colleagues).[57,58] In MeroX, we searched each of our Raw files separately, filtering to a 1% FDR at the XSM level for each search. Validated peptides were then compiled together (Tables S5−S10 show the number of XSMs and identifications in each fraction). In Xi, we examined several distinct search settings because of the greater processing and memory resources this software needed for this large search space and our desire to streamline the computational workflow (Supporting Information, Table S11). With Xi, we concatenated the pre-validated search results from all the fractions that corresponded to the same experiment (see Figure S13 for a flow chart) and performed validation on these in XiFDR with a 5% FDR cutoff at the residue pair level. Once again, several settings were tested, and we found that prefiltering for doublet counts (CCfragmentdoubletcount) to be greater than 0 and boosting on "between" (interprotein) cross-links were beneficial (Supporting Information, Table S12). Hence, we applied these XiFDR settings to all of our experiments; Table S13 (Supporting Information) reports the number of identifications for each experiment searched with either inclusive or conservative settings. In general, the inclusive settings resulted in significantly more cross-links being identified.

To check if these cross-links were reliable, we decided to focus on one of the most structurally well-characterized complexes in *E. coli*, namely, the ribosome (Figure 2B−D). Both MeroX and Xi performed well by this test. MeroX identified 230 cross-links involving only ribosomal proteins, while Xi identified 511; however, most of these were intraprotein cross-links (i.e., between two peptides on the same ribosomal protein), some were interprotein, and these were equally structurally plausible (Figure 2B,C). For each cross-linked peptide identified in MeroX and Xi, we measured the Cα−Cα distance between the cross-linked residues (as based on one cryo-EM structure of the *E. coli* ribosome, PDB: 7CPJ), and Figure 2D shows the cumulative frequency distributions for these distances. Remarkably, 75% of the cross-links have Cα-Cα distances below 15 Å, representing "zero-length" cross-links (given that diazirines will be ∼10 Å away from the Cα of the lysine to which it is attached, and most other residues' side chains extend ∼5 Å from their Cα). The data set was also found to be of high quality using the fraction of structure-corroborating identifications (FSI) metric introduced by Yugandhar et al. (Table S14).[59] The identification of many short-distance cross-links within and between ribosomal proteins highlights an exciting advantage of photo-cross-linkers to provide high-resolution structural information and also suggests that accurate identifications from whole proteomes are possible, at least for ribosomal proteins. Figure 2F shows an inter-protein cross-link between the N-terminus of RplU and D102 of RplT. The assignment of D102 as the attachment site is supported by 15 y-ions that contain the modification and is moreover consistent with the structure (these residues have a Cα−Cα distances of 10.5 Å).

In addition to finding cross-links between ribosomal proteins, we also identified two cross-links between ribosomal proteins (RplL and RplB) and elongation factor EF-G, responsible for translocation (Figure 2E). These two cross-links were also resolved down to the residue-level and cohered closely to a cryo-EM structure (PDB: 7N2C) with the elongation factor bound to the A-site (with Cα−Cα distances of 9.2 and 17.7 Å).

Xi Workflow Better Controls for False Discovery. Because we obtained poorer quality results in our experiments in which cross-linking was conducted in lysate (see Figure S14 for summary data across search algorithms), we focused on the conditions in which cross-linking was conducted *in cellulo*. The number of unique cross-links pooled together from the 16 and 15 respective fractions of the DizSEC and DizSPC cross-linking experiment is shown in Figure 3A, using either MeroX or Xi workflows (green or orange, respectively). In general, Xi identified more intraprotein cross-links for both cross-linkers. Intraprotein cross-links provide structural information; however, interprotein cross-links are generally considered more interesting from the perspective of building PPI networks and assisting molecular docking.[60−62] We were initially surprised to see that MeroX identified significantly (∼3-fold, Figure 3A) more interprotein cross-links than Xi, and this made us wonder if perhaps XiFDR was stricter in its validation criteria for interprotein cross-links. Manually inspecting MeroX's identifications, we found it interesting that so many of the cross-links involved low-abundance or uncharacterized proteins, including those that were not even identified in standard linear searches.

To interrogate this more quantitatively, we categorized *E. coli* proteins as high-abundance (>200 copies/cell), low-abundance (<200 copies/cell), or unknown abundance (as based on Ribo-seq data on *E. coli*[63] documented in EcoCyc[64,65]) and sorted each cross-link into categories based on the abundances of the two proteins it connects. The results are shown in the frequency tables in Figure 3B. Consistent with our suspicion, MeroX identifies many cross-links to low-abundance proteins, but this curiosity is not the case using the Xi workflow with validation by XiFDR. As a further test, we performed an entrapment search on one of our fractions (DizSEC, *in cellulo*, fraction 13) in MeroX with a FASTA file in which half of the records are impossible (taken from the yeast

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

proteome rather than the *E. coli* proteome). Consistent with the other observations, incorrect identifications involving yeast peptides were rather common among interprotein cross-links (Figure 3C). We performed an entrapment search using Xi on the DizSEC in-cell fractions and identified far fewer erroneous yeast cross-links.

On the balance, these further tests lead us to conclude that at present, the Xi workflow provides more reliable identifications of inter-protein cross-links in proteome-wide photo-cross-linking experiments (see Data S1 for compiled Xi cross-links and Figure S15 for a network diagram of *in cellulo* Xi cross-linking data). This finding is broadly consistent with the view that FDR estimation for inter-protein cross-links is particularly challenging, requiring a careful propagation of error from XSMs to peptide pairs to protein pairs.[3]

Diazirines Can Efficiently Engage Non-acidic Residues. Iacobucci and co-workers reported the finding that activated diazirines can react with carboxylic acids (as on Asp and Glu residues) to form MS-cleavable ester linkages.[43] By instructing the search algorithm to look for these cleavage products, cross-links were more readily identified using the diazirine-containing photo-amino acids (photo-Met and photo-Leu), and it was suggested that this may be the dominant mode (ca. 75%) for diazirine cross-link formation. Separate studies in which diazirine photoaffinity labels were combined with individual amino acids have given support to the view that alkyl diazirines preferentially react with acidic residues particularly in the protonated state.[32] On the other hand, since MS-cleavable ester cross-links result in MS^2 spectra that are more easily identified, the notion that diazirines preferentially react with acidic residues might be biased. Because the doubly fissile urea linkages formed by DizSEC and DizSPC enable reliable identification of cross-linked peptides regardless of the photoproduct's insertion site, these compounds could enable an unbiased test of this hypothesis. Figure 3D shows the distribution for the non-lysine attachment site for the DizSEC proteome-wide *in cellulo* cross-linking experiment, which is highly non-random as based on a *P*-value of 10^{-239} according to the Chi-square test.[66] Glu is found to be the most frequent cross-link site (22.5%), a four-fold enrichment above its frequency in the *E. coli* proteome. This is consistent with the theory that diazirines preferentially react with protonated carboxylic acids since Glu is less acidic than Asp, and Asp has a comparatively modest enrichment factor at 1.8-fold. Between the two acidic residues, they account for 32% of all attachment sites from the photo-cross-link (33% for DizSPC). However, in contrast with the view that acidic residues are the universal targets, we find the second most frequent insertion site is Ala (13.2%), which is 1.3-fold enriched above its frequency in the *E. coli* proteome (and 1.7-fold enriched once acidic sites have been discounted). Once acidic residues have been discounted, the amino acid types that are most enriched (relative to the proteomic frequency) for serving as attachment sites are Ala, Gly, Val, and Tyr (in both the DizSEC and DizSPC *in cellulo* experiments, cf. Figure S16). Of note, Gly was a surprisingly frequent cross-linking site (8.3%), despite its lack of a side chain. This is a desirable feature because Gly is common in disordered regions and flexible loops. Figures S17, S18 show representative MS^2 spectra involving Gly as the non-lysine attachment site. It is important to note that in cases where fragmentation patterns are incomplete, XiSearch cannot unambiguously call the placement of the non-lysine attachment site, and so the cross-linking preferences shown in Figure 3D should be viewed with some caution. Nevertheless, such non-random distribution of attachment sites with respect to overall proteomic frequencies with acidic residues discounted (*P*-value 10^{-45} by the Chi-square test) indicates that XiSearch can call attachment sites with some fidelity. Finally, we observed that for 10% of the identified cross-linked peptides, there was no acidic residue present on the photo-cross-linked (non-lysine) side, providing further evidence for participation of non-acidic residues.

Even though DizSEC cross-links frequently with Asp and Glu, there is another subtle reason why the doubly fissile urea fragmentation is an important improvement for reliable identification of photo-cross-links proteome-wide. As shown in Figure 3E, the ester linkage between a conventional diazirine (e.g., photo-methionine) and Glu will cleave to form an alkene and unmodified Glu. Since alkenes are also a potential quench ("dead-end") product from diazirine photolysis, the precursor mass/MS^2 spectrum from a cleaved ester-derived cross-link would be indistinguishable from a scenario in which a peptide with a quenched diazirine and a peptide with an Asp/Glu form an association complex in the gas phase (Figure 3E). This ambiguity is not present when the photo-cross-linker is incorporated onto a doubly fissile MS-cleavable system. Figure 3F shows a representative MS^2 spectrum in which the non-lysine cross-linking site is Glu. Fragmentation at the urea generates the previously discussed modified linear peptides arranged into doublets [lys_{CO} and lys_{pep} (found in two charge states); non_{amine} and $non_{isocyanate}$]. Fragmentation at the ester results in a distinct product, an unmodified linear form of the peptide with the non-lysine attachment site (would be called non_{pep} in our nomenclature scheme). As shown in Figure 3F, the non_{pep} ion is present—consistent with the well-known fact that esters are labile functional groups under HCD. However, if esters were much *more* labile than ureas, then the amine-bearing and isocyanate-bearing modifications on the non-lysine peptide would be rare. This is apparently not the case as the intensity of the fragmentation-at-urea products is comparable to the fragmentation-at-ester product (Figure 3F). Manual interrogation of many spectra showed us that this feature is common among MS^2 spectra in which DizSEC cross-linked at Asp or Glu.

Structural Validation from Heterocomplexes. Intermolecular cross-links are generally more interesting but also more challenging to reliably identify in the enormous search space encompassed by proteome-wide photo-cross-linking.[3] Structurally well-defined heterocomplexes should have restricted cross-link distances if inter-protein cross-linked peptides are identified properly. Therefore, we mapped our cross-links against the reported structures of hetercomplexes in order to obtain a more visual form of validation, although some reports have suggested that other validation strategies are more stringent.[59] For our "final" data set on PPIs, we performed XiSearch/XiFDR on fractions from DizSEC and DizSPC-treated cells, cross-linked *in cellulo*. The samples in which photo-cross-linking was conducted in lysate generated very few inter-protein cross-links when searched and validated in the Xi workflow (Figure S14).

In addition to the ribosome (generally one of the easier targets for XL-MS studies), we also obtained a dense network of cross-links interconnecting the heterocomplex between EF-Tu and EF-Ts, a GTPase associated with translation elongation and its associated guanosine exchange factor (Figure 4A).

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

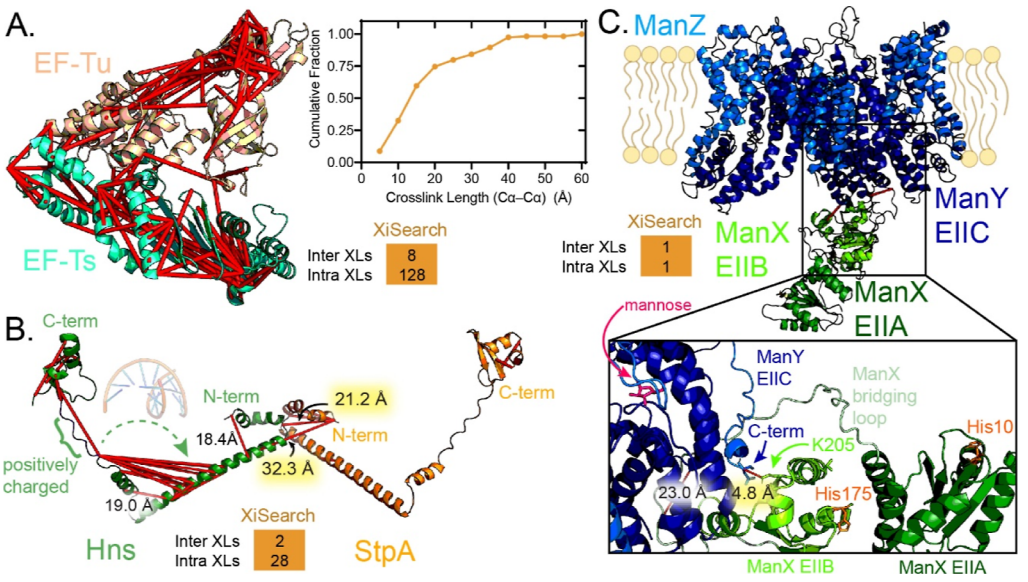


Figure 4. Cross-link validation by comparison to structures of heterocomplexes. (A) Identified cross-links from XiSearch are mapped onto the EF-Tu/EF−Ts complex (PDB: 4PC3). (B) Identified cross-links from XiSearch are mapped onto the H-NS (AlphaFold: P0ACG)/StpA (AlphaFold: P0ACF8) complex, docked in HADDOCK. The N-termini of the proteins interact with one another, while the C-terminal domain and linker interact with DNA. (C) Structural model generated, in which Alphafold structures of ManX (AlphaFold: P69797) and ManY (AlphaFold: P69801) are docked based on cross-links, and then, ManY aligned with the cryo-EM structure (PDB: 6K1H) of the whole pore ($ManY_3ManZ_3$). The mannose bound to loops in ManY and ManZ is shown in magenta, and catalytic phosphohistidines in ManX are shown in orange. Cross-links used for docking are highlighted with yellow background.

Superposing these cross-links onto the X-ray structure of the complex, we once again detect only structurally reasonable cross-links with a median Cα−Cα distance of 15.7 Å.[67]

As another example shown in Figure 4B, we obtained a decent number of cross-links between Hns (histone-like nucleoid structuring protein) and its paralogous partner, StpA (sometimes also called HnsB) which had previously been shown to interact.[68] These proteins promote compaction of DNA and transcriptional repression in response to stress. Using our cross-links alone, we were able to obtain a reasonable dimer structure docking these two proteins' AlphaFold structural models with HADDOCK, which recapitulates the known function of the N-terminal domain in oligomerization.[69,70] The cross-links on Hns also show contacts between the long N-terminal helix and an ostensibly distal C-terminal disordered region, implying an apparent pincer movement about the loop which forms a joint at the bottom of the "V." This is consistent with previous observations on mutants that assigned an important functional role of the linker region in gene repression.

One advantage of cross-linking-based structural approaches is that they are amenable to interrogating disordered regions even though these portions of proteins are typically poorly resolved in X-ray or cryo-EM structures. We obtained one cross-link bridging ManX and ManY and two components of the phosphotransferase system (PTS) responsible for the active uptake of mannose (Figure 4C). ManX comprises two domains that correspond to EIIA and EIIB; the EIIA domain accepts phosphate from a phosphocarrier protein at His10 and then transfers it to His175 on the EIIB domain.[71] The EIIB domain docks with EIIC (corresponding to ManY for the mannose PTS), the transmembrane channel which imports incoming mannose and transfers phosphate to it.[72] The channel is itself a hexamer comprising three copies of ManY and three copies of ManZ. Although the structures of the EIIAB protein (ManX) and the channel (ManYZ) have both been solved,[71,72] the manner in which ManX docks and donates phosphate to mannose has not been seen in structures even though the interaction is supported by genetics and biochemical evidence.[73] To visualize this interaction, we docked ManX to ManY using our cross-linking data with AlphaFold structures (which contain all disordered regions) and found that EIIB on ManX interacts with the disordered cytoplasmic-facing C-terminus of ManY. Hence, in-cell, cross-linking can highlight the importance of disordered regions in promoting a biochemical interaction that could be challenging to resolve.

Insights into Cellular Biology from Proteome-Wide Photo-Cross-Linking. We pooled data from two in-cell experimental cross-linking conditions (DizSEC and DizSPC, both with Lys-C pre-digestion, rows 1 and 3 from Table S13, Supporting Information) and obtained 3202 XSMs and 2362 unique cross-links (257 interprotein and 2105 intraprotein) representing 177 different PPIs. The result is a small portion of *E. coli*'s interaction network, albeit with residue-level resolution. Currently, the coverage from this data set is modest relative to the total *E. coli* interactome, and we would like to point out here that other interactomic methods (such as by high-throughput affinity-purification mass-spectrometry[74] and yeast two-hybrid screens[75]) have to date provided deeper coverage than the present one. Nevertheless, probing interactions in their native cellular context and with residue level-resolution are interesting advantages afforded by proteome-wide photo-cross-linking. Hence, we interrogated this network and focused on a few hotspots where there were denser sets of cross-links available, and in the following describe three examples that highlight how the results from photo-cross-linking can provide new insights at the intersection of cellular and structural biology.

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

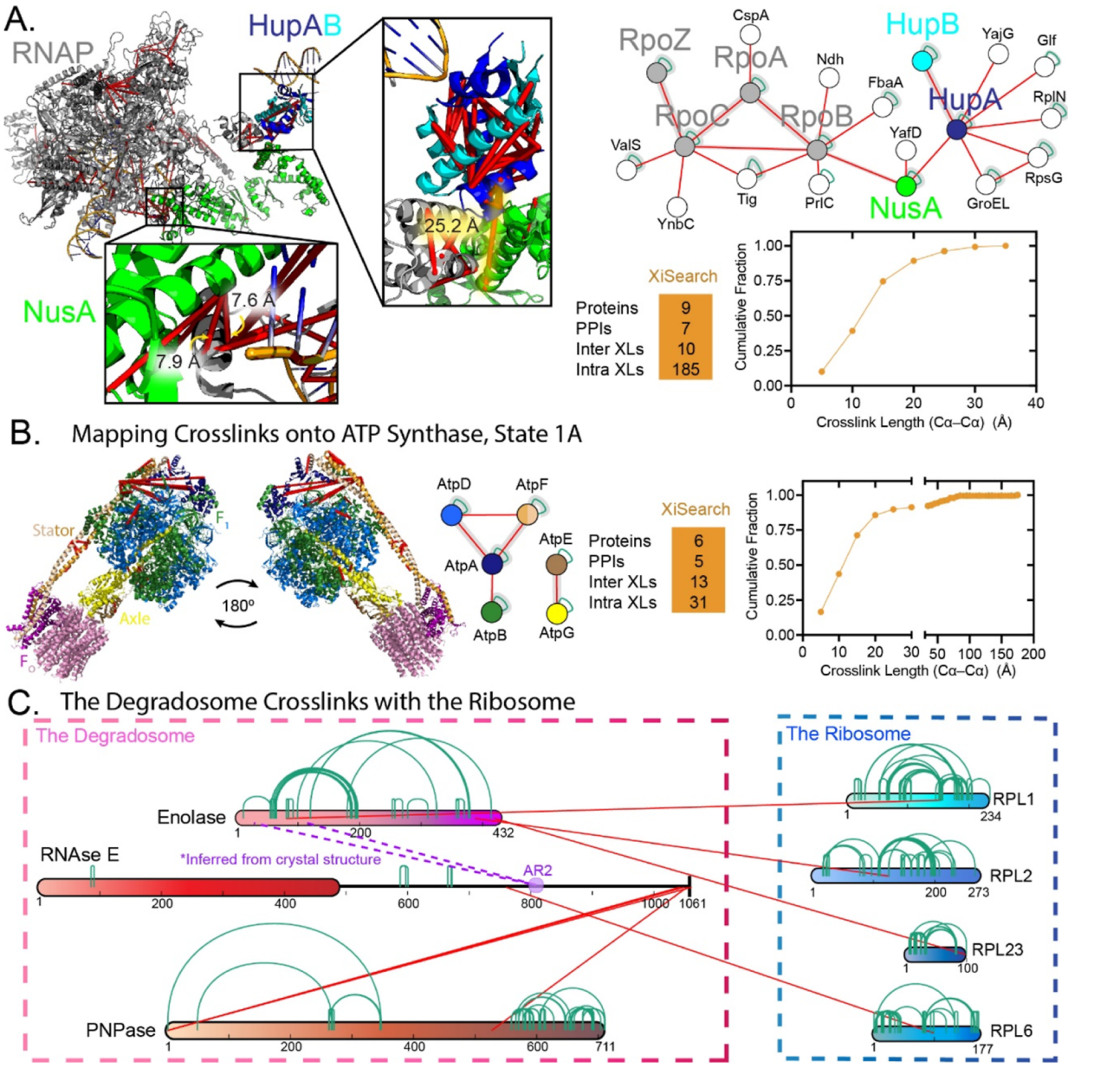


Figure 5. Subnetworks among dynamic complexes reveal structural insights. (A) Subnetwork featuring known interactions between RNAP and NusA (PDB: 6FLQ), and an unreported interaction between NusA and HupA, a histone-like protein in *E. coli*. A structural model in which HupAB (PDB: 4YEW) is docked and cumulative distribution plots of Cα−Cα distances are provided. The cross-link used for docking is highlighted with yellow background. (B) Structural model, network plot, and cumulative distribution plot for the ATP synthase complex (PDB: 6OQR). (C) Subnetwork associated with the *E. coli* degradosome, featuring known interactions between PNPase, RNAse E, and the ribosome. Unreported interaction between enolase and the ribosome suggests a potential role as an adaptor.

One example focuses on a sub-network centering on RNAP (Figure 5A), whose core enzyme in *E. coli* is a tetramer, [$(RpoA)_2$(RpoB)(RpoC)]. The XL-MS data recapitulate the key interactions that make up the core enzyme, with cross-links bridging between all three polypeptides. We further identified a linkage between RpoC and RpoZ, the ω-subunit of RNAP, which is an assembly factor that has been shown to facilitate the incorporation of RpoC into the holo-complex.[76] As before, 75% of the cross-links approach zero-length, highlighting the close-range of these cross-links.

We also detected two cross-links between RNAP and the protein NusA (consistent with a cryo-EM structure of the complex, as shown in left inset in Figure 5A), which has been implicated in stabilizing transcriptional pauses and anti-termination in *E. coli*.[77] Our cross-links show that the N-terminal helix of NusA binds against a hairpin between the first two alpha helices of HupA (right inset in Figure 5A), a subunit of the HU histone-like DNA-binding complex. To visualize this interaction, we used these data to dock HupAB onto the RNAP:NusA complex. HU is a heterodimer containing one monomer of HupA and HupB and is implicated in nucleoid organization and regulation.[78] Because XL-MS data provide spatial resolution for these linkage sites, this cross-link could be used as a distance restraint for future studies focusing on the structural basis of transcriptional regulation in *E. coli*. Moreover, these observations show that the role of NusA is a multifaceted adaptor protein, involved in bridging RNAP to both termination factors and to nucleoid proteins.

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

A second example concerns ATP synthase, another abundant macromolecular complex in *E. coli* responsible for converting proton gradients across the inner membrane into ATP synthesis (Figure 5B). In total, 44 cross-links were observed, including some non-obvious intermolecular cross-links spanning the F_1 subunit and the rotor (AtpA−AtpB). While the majority of the cross-links (75%) are under 15 Å, in this case we observe a "fat tail" to the Cα−Cα distance distribution, with 10% of the cross-links exceeding 50 Å in the cryo-EM structure of state 1A. All these long-distance cross-links occurred between the stator (AtpF) and apical portions of the F_1 subunit (comprising AtpA and AtpD). While difficult to reconcile with a single ATP synthase complex (including in its alternative rotational states[79]), these cross-links could make sense in the context of the "rows" of ATP synthases that have been resolved in cryo-ET reconstructions on yeast mitochondria, in which F1 subunits come very close to one another.[80] Although this formation has not been previously shown in *E. coli*, these results are suggestive that higher order organization of ATP synthases may exist in bacteria as well.

As a final vignette, we were intrigued by a set of cross-links that interconnect RNase E, polynucleotide phosphorylase (PNPase), enolase, and the large subunit of the ribosome (Figure 5C). The degradosome is a large multi-enzyme complex consisting of RNase E, PNPase, enolase, and an RNA helicase (RhlB—not in our observed subnetwork) responsible for mRNA decay in *E. coli*.[81] RNase E has been described as the scaffold for the degradosome, particularly its disordered C-terminal half. Indeed, our cross-links recapitulate a known interaction between the C-terminus of RNase E and PNPase.[82,83] A report from Tsai et al. showed that the degradosome associates stably with ribosomes, which might coordinate the degradation of mRNA following its translation.[84,85] They found that the C-terminal portion of RNase E binds to ribosomes in vitro through electrophoretic mobility shift assays through a putative RNA-binding domain termed the AR2 region (residues 796−819). Consistent with this result, we identified a cross-link between RPL6 and a site in RNase E adjacent to the AR2 region. In addition, we also found several cross-links between enolase and the ribosome, specifically at RPL1, RPL2, and RPL23. These three ribosomal proteins face the E-site of the ribosome; hence, enolase could serve as an adaptor that helps orientate RNase E to the outgoing side of the mRNA channel. This positioning would facilitate degradation of the 5′ side of the mRNA as it originates from the trailing end of the polysome, a model initially proposed by Luisi and colleagues.[84]

An important caveat to all these discussions is that the cross-link density in our data set of any one of the three complexes discussed (RNAP, ATP synthase, and degradosome) is not high enough to develop fully fledged models. It would require more detailed cross-link maps of the individual complexes, likely achieved through tagging and enrichment by affinity purification. Nevertheless, we hope to convey that proteome-wide photo-cross-linking can be a powerful hypothesis generation tool, revealing structural relationships between proteins in situ to stimulate further study.

DISCUSSION AND FUTURE DIRECTIONS

Diazirines have achieved a privileged position in chemical biology because of their versatility as a "universal" cross-linking moiety, reflected in their frequent use as a photoaffinity label in chemoproteomic studies.[86,87] Nevertheless, transferring this versatility to XL-MS has been challenging. Even though diazirine-containing cross-linkers including SDA, sulfo-SDA, and photo-amino acids are commercially available, it has been difficult to apply them to proteome-wide samples because the number of different photo-cross-linked peptides generated is so great and the sequence space against which candidate spectra must be searched even greater.

To tame this challenge, we have introduced DizSEC and DizSPC, diazirine-containing NHS-carbamates which react with lysines to generate MS-cleavable urea linkages. These cross-links are doubly fissile under HCD, generating characteristic modifications and doublets that can assist spectral assignment and placement of the photo-cross-link site. They have enabled us to confidently assign cross-linked spectra in very large search spaces (i.e., wherein any lysine in any *E. coli* protein can cross-link to any other residue in the *E. coli* proteome), including those in which the diazirine reacts with non-acidic residues. While conceptually similar to the previously described SDASO cross-linker (Figure 1A),[41] two potential advantages of DizSEC/DizSPC are the doubly fissile nature of their linkages (assisting spectral search) and their smaller footprints after attachment to lysine (thereby interfering less with protein interactions).

Search software is a major consideration to the workflow for photo-cross-linking, and in this study, we examined two packages, MeroX and XiSearch/XiFDR. We found that the risk for false discovery in the large search space spanned by proteome-wide photo-cross-linking is acute, and currently this challenge seems to be better addressed by the improvements in XiFDR introduced in v2.0.[3] On the other hand, an updated version of XiSearch with better memory allocation management (garbage collection) would be helpful for making these searches more accessible to users who lack access to high-performance computation clusters.

We are excited about the potential for these types of global photo-cross-linking methods. The finding that DizSEC/SPC are cell-permeable and can be rapidly activated *in cellulo* (in 1 s) with our setup suggests that they can be useful tools for probing protein structures and interactions at residue-level resolution in their native context (and with high temporal resolution, though we did not explicitly take advantage of that in this study). Recent years have seen increased interest in characterizing the quinary structure of proteins,[88,89] which refers to the weak but specific interactions that cause the cytosol to demix and enable certain sets of proteins to associate with one another with preferential orientations. These interactions are important for the formation of metabolons and biomolecular condensates. However, there are few techniques that can directly probe protein interaction networks at high structural resolution inside living cells. Currently, XL-MS and cryo-ET are leading contenders with complementary strengths. Cryo-ET studies have typically focused on large molecular complexes with distinct shapes (such as the ribosome or proteasome). Although the spatial resolution of cryo-ET is remarkable and improving, its "compositional resolution"—that is, its ability to identify which specific protein an arbitrary segment of electron density corresponds to—is more limited. XL-MS excels at resolving molecular identity, and hence close collaboration between them is set to enable "molecular sociology" to characterize smaller proteins that are less easy to identify based on electron density alone. In this vein, photo-activated cross-linkers are germane for

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685

providing greater structural and temporal resolution than other XL-MS modalities.

On the other hand, there are still major challenges ahead. To be blunt, the number of intermolecular cross-links we identified (257) from 35 fractions (88 instrument hours) is not enviable and is modest compared to studies that have used conventional chemical cross-linkers[3,4,90−92] or alternative methods.[2,74,75,93,94] Hence, further improvements to the workflow will be needed to make photo-cross-linking deliver on these high expectation; we conclude with where we suspect progress could be made.

First, given the complexity of photo-cross-linked proteomes, it is likely that even greater fractionation than we employed here would be necessary to mitigate co-isolation problems.[95] A likely candidate is 2D fractionation combining SEC (as done here) with strong-cation exchange, which has proven successful for other high-complexity cross-linking experiments.[96] Second, our experiments would have likely benefitted from tighter instrument calibration and higher resolution on MS^2 scans. XiFDR is well-equipped at filtering out decoys, but this comes with the requirement that candidate spectra match their targets with high precision. Future experiments will need to be more exacting in this choice of parameters. Finally, studies that employ affinity purification or organelle isolation would likely be preferrable to undifferentiated lysates.[95] Although the ability to search a proteome-wide database with photo-cross-linkers is an important milestone demonstrated here, undifferentiated lysates typically overrepresent high-abundance proteins (such as the ribosome) and are likely not the best path forward to delve deeper into the proteome's dynamic range.[95]

To summarize, we present new molecular tools, coupled with an experimental and computational workflow, that enable reliable proteome-wide identification of photo-cross-linked peptides. This study has piloted these tools to produce a small, partial PPI network of *E. coli* but nevertheless one that is resolved down to the residue-level. We have performed comparisons to known structures of protein complexes and verified short Cα−Cα distances to build confidence in the spectral identifications and have used the data set as a hypothesis-generating tool to propose three potentially new functional interactions associated with RNAP, ATP synthase, and the degradosome. With the proposed improvements suggested in this final section, we hope photo-XL-MS will soon contribute meaningfully to our understanding of molecular sociology inside cells.

■ ASSOCIATED CONTENT

Data Availability Statement

Raw data and settings files corresponding to the in-cell cross-linking experiments (31 LC−MS/MS runs in total) have been deposited to PRIDE under accession code PXD041148. Data processing scripts are available under the Zenodo deposition (10.5281/zenodo.7778901).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c01369.

All the additional data of cross-linking experiments, synthetic procedures, and characterization of synthesized compounds (PDF)

Pooled unique cross-links identified in the lysate and in-cell cross-linking experiments (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Stephen D. Fried − *Department of Chemistry and Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0003-2494-2193; Email: sdfried@jhu.edu

Authors

Anneliese M. Faustino − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Piyoosh Sharma − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0002-1102-8602

Edgar Manriquez-Sandoval − *Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Divya Yadav − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c01369

Author Contributions

§A.M.F. and P.S. contributed equally to this study.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.D.F. acknowledges support from the NIH Director's New Innovator Award (DP2GM140926) and an HFSP research grant (RGY0074/2019). P.S. would like to acknowledge support from the Albstein Foundation for Brain Research. The authors would like to thank Yingzi Xia for SEC method development and Phil Mortimer for providing the critical mass spectral support. The authors would like to give special thanks to Tyrel McQueen for providing access to the Advanced Research Computing at Hopkins (ARCH) core facility (rockfish.jhu.edu), which is supported by the National Science Foundation (NSF) grant number OAC1920103. The authors are especially grateful to Andrea Graziadei and Lutz Fischer for technical support with utilizing Xi software suite and for providing data processing scripts. We thank Francis O'Reilly for thoughtful and helpful conversations.

■ REFERENCES

(1) Iacobucci, C.; Götze, M.; Sinz, A. *Curr. Opin. Biotechnol.* **2020**, *63*, 48−53.

(2) Mishra, P. K.; Yoo, C.; Hong, E.; Rhee, H. W. *ChemBioChem* **2020**, *21*, 924−932.

(3) Lenz, S.; Sinn, L. R.; O'Reilly, F. J.; Fischer, L.; Wegner, F.; Rappsilber, J. *Nat. Commun.* **2021**, *12*, 3564−3611.

(4) Wheat, A.; Yu, C.; Wang, X.; Burke, A. M.; Chemmama, I. E.; Kaake, R. M.; Baker, P.; Rychnovsky, S. D.; Yang, J.; Huang, L. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2023360118.

(5) O'Reilly, F. J.; Xue, L.; Graziadei, A.; Sinn, L.; Lenz, S.; Tegunov, D.; Blötz, C.; Singh, N.; Hagen, W. J. H.; Cramer, P.; et al. *Science* **2020**, *369*, 554−557.

(6) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell. Proteomics* **2010**, *9*, 1634−1649.

(7) Sinz, A. *J. Mass Spectrom.* **2003**, *38*, 1225−1237.

(8) Trakselis, M. A.; Alley, S. C.; Ishmael, F. T. *Bioconjugate Chem.* **2005**, *16*, 741−750.

(9) Bishop, A. C.; Buzko, O.; Shokat, K. M. *Trends Cell Biol.* **2001**, *11*, 167−172.

(10) Petrotchenko, E. v.; Borchers, C. H. *Mass Spectrom. Rev.* **2010**, *29*, 862−876.
(11) Tang, X.; Bruce, J. E. *Mol. BioSyst.* **2010**, *6*, 939−947.
(12) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. *Nat. Methods* **2015**, *12*, 1179−1184.
(13) Arora, B.; Tandon, R.; Attri, P.; Bhatia, R. *Curr. Protein Pept. Sci.* **2017**, *18*, 946−955.
(14) Matzinger, M.; Mechtler, K. *J. Proteome Res.* **2020**, *20*, 78−93.
(15) O'Reilly, F. J.; Rappsilber, J. *Nat. Struct. Mol. Biol.* **2018**, *25*, 1000−1008.
(16) Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. *Anal. Chem.* **2019**, *91*, 10236−10244.
(17) 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.; et al. *J. Proteome Res.* **2017**, *16*, 2457−2471.
(18) Liu, F.; Lössl, P.; Scheltema, R.; Viner, R.; Heck, A. J. R. *Nat. Commun.* **2017**, *8*, 15473.
(19) Fürsch, J.; Kammer, K.-M.; Kreft, S. G.; Beck, M.; Stengel, F. *Anal. Chem.* **2020**, *92*, 4016−4022.
(20) Xue, L.; Lenz, S.; Zimmermann-Kogadeeva, M.; Tegunov, D.; Cramer, P.; Bork, P.; Rappsilber, J.; Mahamid, J. *Nature* **2022**, *610*, 205−211.
(21) Schweppe, D. K.; Chavez, J. D.; Lee, C. F.; Caudal, A.; Kruse, S. E.; Stuppard, R.; Marcinek, D. J.; Shadel, G. S.; Tian, R.; Bruce, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 1732−1737.
(22) Chavez, J. D.; Lee, C. F.; Caudal, A.; Keller, A.; Tian, R.; Bruce, J. E. *Cell Syst.* **2018**, *6*, 136−141.
(23) Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. *Nat. Protoc.* **2018**, *13*, 2964−2990.
(24) Kaake, R. M.; Wang, X.; Burke, A.; Yu, C.; Kandur, W.; Yang, Y.; Novtisky, E. J.; Second, T.; Duan, J.; Kao, A.; et al. *Mol. Cell. Proteomics* **2014**, *13*, 3533−3543.
(25) Jiao, F.; Yu, C.; Wheat, A.; Wang, X.; Rychnovsky, S. D.; Huang, L. *Anal. Chem.* **2022**, *94*, 4236−4242.
(26) Moss, R. A. *Acc. Chem. Res.* **2006**, *39*, 267−272.
(27) Das, J. *Chem. Rev.* **2011**, *111*, 4405−4417.
(28) Deng, H.; Lei, Q.; Wu, Y.; He, Y.; Li, W. *Eur. J. Med. Chem.* **2020**, *191*, 112151.
(29) Hu, W.; Yuan, Y.; Wang, C.-H.; Tian, H.-T.; Guo, A.-D.; Nie, H.-J.; Hu, H.; Tan, M.; Tang, Z.; Chen, X.-H. *Chem* **2019**, *5*, 2955−2968.
(30) Dubinsky, L.; Krom, B. P.; Meijler, M. M. *Bioorg. Med. Chem.* **2012**, *20*, 554−570.
(31) Zhuo, F.; Guo, Q.; Zheng, Y.; Liu, T.; Yang, Z.; Xu, Q.; Jiang, Y.; Liu, D.; Zeng, K.; Tu, P. *ChemBioChem* **2022**, *23*, No. e202200038.
(32) West, A. V.; Muncipinto, G.; Wu, H.-Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. *J. Am. Chem. Soc.* **2021**, *143*, 6691−6700.
(33) Halloran, M. W.; Lumb, J. *Chem.—Eur. J.* **2019**, *25*, 4885−4898.
(34) Yang, T.; Liu, Z.; Li, X. D. *Chem. Sci.* **2015**, *6*, 1011−1017.
(35) Yang, T.; Li, X.-M.; Bao, X.; Fung, Y. M. E.; Li, X. D. *Nat. Chem. Biol.* **2016**, *12*, 70−72.
(36) Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P. R. *Nat. Commun.* **2016**, *7*, 12299−12310.
(37) Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R. *J. Am. Chem. Soc.* **2014**, *136*, 11860−11863.
(38) Kalkhof, S.; Sinz, A. *Anal. Bioanal. Chem.* **2008**, *392*, 305−312.
(39) Wang, J.; Kubicki, J.; Peng, H.; Platz, M. S. *J. Am. Chem. Soc.* **2008**, *130*, 6604−6609.
(40) Blencowe, A.; Hayes, W. *Soft Matter* **2005**, *1*, 178−205.
(41) Gutierrez, C.; Salituro, L. J.; Yu, C.; Wang, X.; DePeter, S. F.; Rychnovsky, S. D.; Huang, L. *Mol. Cell. Proteomics* **2021**, *20*, 100084.
(42) Häupl, B.; Ihling, C. H.; Sinz, A. *Proteomics* **2017**, *17*, 1600459.
(43) Iacobucci, C.; Götze, M.; Piotrowski, C.; Arlt, C.; Rehkamp, A.; Ihling, C.; Hage, C.; Sinz, A. *Anal. Chem.* **2018**, *90*, 2805−2809.
(44) Yu, C.; Novitsky, E. J.; Cheng, N. W.; Rychnovsky, S. D.; Huang, L. *Anal. Chem.* **2020**, *92*, 6026−6033.
(45) 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.
(46) Mahamid, J.; Pfeffer, S.; Schaffer, M.; Villa, E.; Danev, R.; Kuhn Cuellar, L.; Förster, F.; Hyman, A. A.; Plitzko, J. M.; Baumeister, W. *Science* **2016**, *351*, 969−972.
(47) Pfeffer, S.; Mahamid, J. *Curr. Opin. Struct. Biol.* **2018**, *52*, 111−118.
(48) Freer, R.; McKillop, A. *Synth. Commun.* **1996**, *26*, 331−349.
(49) Höfle, G.; Steglich, W.; Vorbrüggen, H. *Angew. Chem., Int. Ed. Engl.* **1978**, *17*, 569−583.
(50) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. *Nat. Biotechnol.* **2012**, *30*, 918−920.
(51) Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. *J. Am. Soc. Mass Spectrom.* **2011**, *23*, 76−87.
(52) Graham, M.; Combe, C.; Kolbowski, L.; Rappsilber, J. *BioRxiv* **2019**, 561829 https://www.biorxiv.org/content/10.1101/561829v1.
(53) Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A. M. J. J. *Front. Mol. Biosci.* **2021**, *8*, 729513.
(54) Van Zundert, G. C. P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; De Vries, S. J.; Bonvin, A. *J. Mol. Biol.* **2016**, *428*, 720−725.
(55) Kolbowski, L.; Lenz, S.; Fischer, L.; Sinn, L. R.; O'Reilly, F. J.; Rappsilber, J. *Anal. Chem.* **2022**, *94*, 7779−7786.
(56) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. *Nat. Protoc.* **2018**, *13*, 2864−2889.
(57) 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.; et al. *Mol. Syst. Biol.* **2019**, *15*, No. e8994.
(58) Graham, M.; Combe, C.; Kolbowski, L.; Rappsilber, J. *bioRxiv* **2019**, 561829.
(59) Yugandhar, K.; Wang, T.-Y.; Wierbowski, S. D.; Shayhidin, E. E.; Yu, H. *Nat. Methods* **2020**, *17*, 985−988.
(60) Graziadei, A.; Rappsilber, J. *Structure* **2022**, *30*, 37−54.
(61) Klykov, O.; Steigenberger, B.; Pektaş, S.; Fasci, D.; Heck, A. J. R.; Scheltema, R. A. *Nat. Protoc.* **2018**, *13*, 2964−2990.
(62) Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. *Chem. Rev.* **2021**, *122*, 7500−7531.
(63) Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S. *Cell* **2014**, *157*, 624−635.
(64) Keseler, I. M.; Mackie, A.; Santos-Zavaleta, A.; Billington, R.; Bonavides-Martínez, C.; Caspi, R.; Fulcher, C.; Gama-Castro, S.; Kothari, A.; Krummenacker, M.; et al. *Nucleic Acids Res.* **2017**, *45*, D543−D550.
(65) Keseler, I. M.; Gama-Castro, S.; Mackie, A.; Billington, R.; Bonavides-Martínez, C.; Caspi, R.; Kothari, A.; Krummenacker, M.; Midford, P. E.; Muñiz-Rascado, L.; et al. *Front. Microbiol.* **2021**, *12*, 711077.
(66) Moura, A.; Savageau, M. A.; Alves, R. *PLoS One* **2013**, *8*, No. e77319.
(67) Kawashima, T.; Berthet-Colominas, C.; Wulff, M.; Cusack, S.; Leberman, R. *Nature* **1996**, *379*, 511−518.
(68) Williams, R. M.; Rimsky, S.; Buc, H. *J. Bacteriol.* **1996**, *178*, 4335−4343.
(69) Rimsky, S. *Curr. Opin. Microbiol.* **2004**, *7*, 109−114.
(70) Dorman, C. J.; Hinton, J. C. D.; Free, A. *Trends Microbiol.* **1999**, *7*, 124−128.
(71) Stolz, B.; Huber, M.; Marković-Housley, Z.; Erni, B. *J. Biol. Chem.* **1993**, *268*, 27094−27099.
(72) Liu, X.; Zeng, J.; Huang, K.; Wang, J. *Cell Res.* **2019**, *29*, 680−682.
(73) Erni, B.; Zanolari, B.; Kocher, H. P. *J. Biol. Chem.* **1987**, *262*, 5238−5247.

(74) Hu, P.; Janga, S. C.; Babu, M.; Díaz-Mejía, J. J.; Butland, G.; Yang, W.; Pogoutse, O.; Guo, X.; Phanse, S.; Wong, P.; et al. *PLoS Biol.* **2009**, *7*, No. e1000096.
(75) Rajagopala, S. V.; Sikorski, P.; Kumar, A.; Mosca, R.; Vlasblom, J.; Arnold, R.; Franca-Koh, J.; Pakala, S. B.; Phanse, S.; Ceol, A.; et al. *Nat. Biotechnol.* **2014**, *32*, 285−290.
(76) Ghosh, P.; Ishihama, A.; Chatterji, D. *Eur. J. Biochem.* **2001**, *268*, 4621−4627.
(77) Guo, X.; Myasnikov, A. G.; Chen, J.; Crucifix, C.; Papai, G.; Takacs, M.; Schultz, P.; Weixlbaumer, A. *Mol. Cell* **2018**, *69*, 816−827.e4.
(78) Kar, S.; Edgar, R.; Adhya, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16397−16402.
(79) Sobti, M.; Walshe, J. L.; Wu, D.; Ishmukhametov, R.; Zeng, Y. C.; Robinson, C. V.; Berry, R. M.; Stewart, A. G. *Nat. Commun.* **2020**, *11*, 2615.
(80) Davies, K. M.; Anselmi, C.; Wittig, I.; Faraldo-Gómez, J. D.; Kühlbrandt, W. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 13602−13607.
(81) Py, B.; Causton, H.; Mudd, E. A.; Higgins, C. F. *Mol. Microbiol.* **1994**, *14*, 717−729.
(82) Kido, M.; Yamanaka, K.; Mitani, T.; Niki, H.; Ogura, T.; Hiraga, S. *J. Bacteriol.* **1996**, *178*, 3917−3925.
(83) Vanzo, N. F.; Li, Y. S.; Py, B.; Blum, E.; Higgins, C. F.; Raynal, L. C.; Krisch, H. M.; Carpousis, A. J. *Genes Dev.* **1998**, *12*, 2770−2781.
(84) Tsai, Y.-C.; Du, D.; Domínguez-Malfavón, L.; Dimastrogiovanni, D.; Cross, J.; Callaghan, A. J.; García-Mena, J.; Luisi, B. F. *Nucleic Acids Res.* **2012**, *40*, 10417−10431.
(85) Aït-Bara, S.; Carpousis, A. J.; Quentin, Y. *Mol. Genet. Genomics* **2015**, *290*, 847−862.
(86) Conway, L. P.; Jadhav, A. M.; Homan, R. A.; Li, W.; Rubiano, J. S.; Hawkins, R.; Lawrence, R. M.; Parker, C. G. *Chem. Sci.* **2021**, *12*, 7839−7847.
(87) Meissner, F.; Geddes-McAlister, J.; Mann, M.; Bantscheff, M. *Nat. Rev. Drug Discovery* **2022**, *21*, 637−654.
(88) Cohen, R. D.; Pielak, G. J. *Protein Sci.* **2017**, *26*, 403−413.
(89) Guin, D.; Gruebele, M. *Chem. Rev.* **2019**, *119*, 10691−10717.
(90) Götze, M.; Iacobucci, C.; Ihling, C. H.; Sinz, A. *Anal. Chem.* **2019**, *91*, 10236−10244.
(91) Gonzalez-Lozano, M. A.; Koopmans, F.; Sullivan, P. F.; Protze, J.; Krause, G.; Verhage, M.; Li, K. W.; Liu, F.; Smit, A. B. *Sci. Adv.* **2020**, *6*, No. eaax5783.
(92) Stieger, C. E.; Doppler, P.; Mechtler, K. *J. Proteome Res.* **2019**, *18*, 1363−1370.
(93) Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. *J. Cell Biol.* **2012**, *196*, 801−810.
(94) Gingras, A.-C.; Abe, K. T.; Raught, B. *Curr. Opin. Chem. Biol.* **2019**, *48*, 44−54.
(95) Lee, K.; O'Reilly, F. J. *Essays Biochem.* **2023**, *67*, 215−228.
(96) O'Reilly, F. J.; Graziadei, A.; Forbrig, C.; Bremenkamp, R.; Charles, K.; Lenz, S.; Elfmann, C.; Fischer, L.; Stülke, J.; Rappsilber, J. *Mol. Syst. Biol.* **2023**, *19*, No. e11544.

https://doi.org/10.1021/acs.analchem.3c01369
Anal. Chem. 2023, 95, 10670−10685