# Isobaric crosslinking mass spectrometry technology for studying conformational and structural changes in proteins and complexes

Jie Luo and Jeff Ranish*

Institute for Systems Biology

401 Terry Avenue North

Seattle, WA 98109

*corresponding author: jeff.ranish@isbscience.org

## Abstract

Dynamic conformational and structural changes in proteins and protein complexes play a central and ubiquitous role in the regulation of protein function, yet it is very challenging to study these changes, especially for large protein complexes, under physiological conditions. Here we introduce a novel isobaric crosslinker, Qlinker, for studying conformational and structural changes in proteins and protein complexes using quantitative crosslinking mass spectrometry (qCLMS). Qlinkers are small and simple, amine-reactive molecules with an optimal extended distance of ~10 Å and use MS2 reporter ions for relative quantification of Qlinker-modified peptides. We synthesized the 2-plex Q2linker and showed that the Q2linker can provide quantitative crosslinking data that pinpoints the key conformational and structural changes in biosensors and the RNA polymerase II (pol II) complex.

## Introduction

Proteins play a central role in the regulation of most biological processes by interacting with other molecules in complexes and interaction networks. Their ability to interact with other molecules to regulate biological processes depends on their structures. Protein tertiary structure is determined by and can be predicted from its primary amino acid sequence[1]. However, proteins, and protein complexes, are not static entities; they can assume multiple conformations. Conformational changes can be induced by post-translational modifications (PTMs), or through interactions with other molecules, and they allow proteins to execute their functions in a condition-dependent manner[2-5]. Many disease-causing mutations impact protein function by altering the conformational landscape of a protein or protein complex[6-8]. Experimental determination of protein conformational states is crucial in many cases to understand the mechanism of protein function and how function is altered in disease. There are two strategies, direct and indirect, for studying protein conformational changes. The direct strategy involves using high resolution methods, such as X-ray crystallography, nuclear magnetic resonance (NMR), small angle X-ray scattering (SAXS), or cryo-EM, to directly observe/compare structural changes under different conditions[9-13]. Aside from the difficulties and limitations associated with each method, additional challenges for studying conformational changes include the fact that not all of the conformations can be easily crystallized or resolved, and many physiological conditions are not compatible with these methods.

The indirect strategies are not able to directly observe the structural changes but can be used to deduce/infer structural changes based on changes in the output of the approach. These methods include limited proteolysis, immunochemical assays, fluorescence resonance energy transfer (FRET), chemical footprinting, and chemical crosslinking, some of which can be performed on proteins/complexes under physiological conditions[14-17]. The main challenge for the indirect strategies is interpreting the data to infer the correct conformational/structural change out of numerous possibilities. Due to the limitations of direct and indirect approaches, the field of protein structural dynamics greatly benefits by combining these strategies with computational modeling[18-20].

Several mass spectrometry (MS)-based methods have been developed as indirect strategies to study conformational and structural changes in proteins and complexes. Hydrogen-Deuterium exchange (HDX) mass spectrometry monitors the isotopic exchange rate between amide hydrogens along the protein backbone and the surrounding solvent to provide information about the folded state of the protein or complex[21]. Protein painting-MS distinguishes solvent accessible regions from inaccessible regions, by using chemical dyes which non-covalently bind to solvent accessible regions, protecting them from trypsin digestion[22, 23]. Protein footprinting uses covalent chemical modifications and MS to identify surface exposed regions. Modification methods include dimethylation of lysine residues by formaldehyde and reducing reagent[24], glycine ethyl ester (GEE)/EDC modification of carboxyl groups on aspartic and glutamic acid[25], NHS-based modification of lysine residues using isobaric Tandem Mass Tag (TMT) labeling reagents[26], isotopic succinic anhydride modification of lysine residues and isotopic N-ethylmaleimide (NEM) modification of cysteines[27]. Hydroxyl radical protein footprinting (HRPF), especially the fast photochemical oxidation of proteins (FPOP), coupled with MS has greatly advanced to permit the study of surface exposed regions of proteins in cell lysates and *in vivo*[28-30]. All of these MS-based methods involve reactions with solvent-accessible residues or surfaces, and are suitable for studying structural changes in proteins and complexes. However, since these methods do not provide information about spatial proximities between residues/domains, information about the relative locations of residues/domains in different conformational states is limited, especially for larger assemblies. Also, a conformational change does not necessarily involve a change in solvent accessibility.

Crosslinking-mass spectrometry (CLMS or CXMS) has been widely used to provide distance restraints between residues and relative positioning of domains in large protein complexes[31-34]. Importantly, when performed in a quantitative manner, CLMS can provide information about conformational changes in proteins and complexes, as well as changes in protein-protein interactions. Multiple quantitative crosslinking-MS (qCLMS) strategies have been developed to study conformational changes. TMT-labeling of crosslinked samples has been used to quantify crosslinked and normal peptides[35, 36]. In this approach, two or more crosslinked samples are individually digested with trypsin and then labeled with TMT reagents prior to combining the samples. The multiple steps prior to combining the samples can result in quantification artifacts that are due to sample processing differences rather than to conformational or interaction changes, thus complicating data interpretation. To avoid these issues, qCLMS strategies that involve co-digestion of crosslinked samples have been developed. These approaches employ isotopically

labelled crosslinkers such as d0/d4 bis[sulfosuccinimidyl] suberate ($BS^3$)[37, 38], d0/d4 bis(sulfosuccinimidyl)-glutarate, -pimelate, and -sebacate[39], d0/d12 ethylene glycol bis(succinimidylsuccinate) (EGS)[40], d0/d12 MS cleavable DSBU[41], d0/d8 MS cleavable CBDPS[42] , or SILAC labelled samples[43]. Relative quantification is based on the MS1 intensities of the isotopically heavy and light crosslinked peptides after identification of the crosslinked peptides. However, like all MS strategies that are based on isotopically heavy and light labeled peptides, the differentially labeled samples double or triple the MS1 complexity, which can decrease sensitivity and reproducibility; many identified crosslinked peptides are difficult to confidently quantify because they are low abundance and difficult distinguish from the background noise signals; and multiple or differential database-searches result in extra difficulties for confident light/heavy crosslinked peptide identification and quantification. Based on their MS-cleavable protein-interaction reporter (PIR) design, Bruce and colleagues have developed 2-plex and 6-plex isobaric quantitative PIRs (iqPIRs) with limited sample handling prior to MS analysis, a single MS1 spectrum for each crosslinker-modified peptide and quantification based on MS2 reporter ions[44, 45]. However, PIR crosslinkers are bulky molecules which are more suitable for studying dynamic protein-protein interactions rather than conformational and structural changes. In fact, many of the aforementioned crosslinker designs are too complicated for routine studies of conformational and structural changes in proteins and complexes[46]. Here we report a novel, simple and small, isobaric crosslinker design for studying conformational and structural changes. This new crosslinker, we termed Qlinker, has an optimal spacer arm length (~10 Å) for crosslinking-based studies of protein conformational and structural changes, and uses 1-imino-2.6-dimethylpiperidin-1-ium reporter ions (similar to TMT reporter ions[47]) for quantification. We have used the 2-plex Q2linker to study conformational changes in biosensors and structural rearrangements that accompany the transition of RNA polymerase II (pol II) from a 10 subunit core complex to a 12 subunit holoenzyme. In each of these studies, the Qlinker approach provided quantitative crosslinking data that pinpointed the key conformational and structural transitions in the proteins/complexes. We expect that this novel strategy for studying conformational and structural changes in proteins and complexes will be easily adopted for routine use by the community.

## Results

### *Experimental procedure and Q2linker quantification*

The two isobaric Q2linkers are named C1q2 and C2q2, synthesized from 1-$^{13}$C and 2-$^{13}$C BAA respectively (Figs. 1a and S1, and Methods). A typical experiment uses equal amounts of C1q2 and C2q2 to crosslink equal amounts of proteins or protein complexes in different structural states. After quenching the reactions with ammonium sulfate, the two samples are combined and mixed well, reduced and denatured, trypsin digested, and analyzed by mass spectrometry using a stepped HCD MS2 method[48]. The Qlinker-modified peptide spectra are identified as either monolinks or crosslinks by different database search engines. The ion signals from the 126.1277 (C1q2) and 127.1311 (C2q2) reporter ions are then extracted from identified spectra to compute the relative abundance of the modified peptides in the different samples (Fig. 1a).

To evaluate the ability of Q2linkers to provide accurate quantification of crosslinker-modified peptides, we crosslinked equal amounts of affinity purified RNA polymerase I (pol I) with either C1q2 or C2q2, digested the two samples separately, and then combined the resulting peptides at known ratios for MS analysis (Fig. S2). Examples of a monolinked peptide spectrum (Fig. 1b) and a crosslinked spectrum (Fig. 1c) are shown with the reporter ion intensities from the indicated mixing ratios. The 126 and 127 reporter ion intensities were extracted from the identified spectra within a mass tolerance of 0.005 Da (~ 40 ppm) and adjusted by correction factors calculated from the isotopic distribution (Methods). The measured log2 ratios of 127/126 intensity were then compared with the expected log2 ratios. The measured ratios from both monolink spectra (Fig. 1d) and crosslink spectra (Fig. 1e) were strongly correlated with the expected ratios with a slope close to 1. At each mixing ratio, the average measured ratio was slightly lower than the expected ratio (intercept ~ -0.4), with a standard deviation around ± 0.25, suggesting there were slightly more crosslinker-modified peptides in the C1q2 sample prior to mixing. This may be due to some peptide loss in the C2q2 sample during trypsin digestion and C18 preparation prior to mixing, or there was slightly less C2q2 crosslinker in the reaction. This also highlights the importance of combining crosslinked samples at an early stage to minimize sample handling variations. The standard deviation is larger for the samples mixed at 5:1 and 10:1 C1q2 labeled (126) to C2q2 labeled (127) (around ± 0.42 and ± 0.78, respectively). We think this could be due to either the higher natural isotope contribution (~10%) from the 126 to the 127 reporter ion or interference from other fragment ions. Overall, these experiments demonstrate that the Q2 quantification strategy can accurately reflect the relative abundance of Q2linker-modified peptides.

We then affinity purified pol II from a yeast strain expressing a FLAG-tagged version of pol II subunit Rpb3 and crosslinked equal amounts of the 12 subunit pol II complex with Q2linkers in the presence and absence of α-amanitin. No major structural rearrangements occur in pol II upon α-amanitin binding (1i3q.pdb for free pol II and 1k83.pdb and 3cqz.pdb for α-amanitin bound pol II). The distributions of log2(127/126) ratios for monolinked (total 174), intralinked (total 154) and interlinked (total 77) peptides were all centered around 0 with no significant differences. 98% of the log2 ratios fell within the range ± 0.5 (Fig. S3). This experiment demonstrates that Q2linkers provide reliable crosslink and monolink quantification with no abnormal ratios when applied to a protein complex in a situation where no significant conformational changes and/or structural rearrangements are known to occur. We initially designed this experiment thinking that α-amanitin binding to pol II might affect Q2linker modification of a lysine residue(s) near the α-amanitin binding site, and we might be able to identify the α-amanitin binding site by analysis of quantitative Q2linker mass spectrometry data. Unfortunately, no Q2linker-modified peptides were identified near the site where α-amanitin binds. This experiment also highlights one of the limitations of residue-specific, quantitative CLMS methods in general. Reactive residues must be available near the region of interest, and the modified peptides must be identifiable by mass spectrometry.

*The study of conformational changes in biosensors using Q2linkers*

Protein biosensors are polypeptides that undergo conformational changes or switches upon receiving an input signal. We next designed experiments to see whether Q2linkers can detect conformational changes in biosensors. We tested two commercially available proteins: maltose binding protein (MBP) and calmodulin (CaM). MBP is a 370 amino acid polypeptide that buries the maltose ligand between a cleft formed by its two domains, and exhibits a conformational change from an open ligand-free conformation (1mpb.pdb) and a closed maltose-bound conformation (1n3w.pdb). However, it is the "balancing interface" on the opposite side of the ligand binding cleft that maintains the open conformation; upon maltose binding, this interface is disrupted and becomes more solvent exposed[49]. To evaluate the ability of Q2linkers to detect conformational changes in MBP, we performed a crosslinker swapping experiment in which we crosslinked 10 ug of MBP in the presence and absence of 10 mM maltose with C1q2 and C2q2, respectively, in one experiment, and, in a second experiment, we crosslinked MBP in the presence and absence of maltose with C2q2 and C1q2, respectively. In each experiment, we combined the C1q2 and C2q2 crosslinked samples prior to trypsin digestion and C18 cleaning, and used 1 ug for mass spectrometry analysis. In both experiments, we identified one monolinked peptide ($_{306}$SYEEELAK*DPR$_{316}$) that was 5-fold more abundant in the samples containing maltose compared to the samples without maltose (Fig. 2a), suggesting K313 is modified more readily by the Q2linkers when maltose is present. Interestingly, this peptide happens to reside within the "balancing interface" of MBP. In the closed conformation, the sequence between 301 and 312 forms an alpha helix (Fig. 2a, gray structure) and upon binding to maltose, the alpha helix relaxes and unwinds (Fig. 2a cyan structure). However, K313 itself is solvent exposed in both structures, so the increase in K313 modification in the open conformation cannot be simply explained by an increase in accessibility upon maltose binding. Comparison of the open and closed conformations of MBP reveals that K313 can form a salt bond with E310 in the open conformation (K313-NE to E310-OE2 distance of ~3.3 Å), and this salt bond is broken in the closed conformation. It is likely that the salt bond between K313 and E310 limits the NHS ester-based modification of K313 as revealed by the quantitative difference in the abundance of the Q2linker-modified peptide containing K313.

Calmodulin (CaM), a small polypeptide composed of 148 AAs, is a well-studied biosensor that undergoes conformational changes upon binding to calcium and CaM-binding peptides (CBPs)[50]. To evaluate the ability of Q2linkers to detect conformational changes in CaM, we crosslinked 25 ug CaM, in the presence or absence of 20 mM $CaCl_2$ and a nearly 1:1 molar ratio of CBP (8 ug) derived from smooth muscle myosin light-chain kinase (MLCK) with C2q2 and C1q2, respectively. After crosslinking, the samples were combined, trypsin digested, and 1 ug of the sample was analyzed by mass spectrometry. One crosslinked peptide was identified containing a Lys78 to Lys95 linkage which was 5-fold less abundant in the presence of CBP and $CaCl_2$ (Fig. 2b), suggesting that crosslinking between these two residues is greatly inhibited upon calcium and CBP association with CaM. We also identified the corresponding monolinked peptides containing these two residues, and found that their abundances were slightly increased upon binding of calcium and CBP (Fig. 2b). These results suggest that the reduced abundance of the crosslinked peptide containing the Lys78-Lys95 linkage cannot be explained by limited accessibility of these two sites

in the presence of $Ca^{2+}$ and CBP. Indeed, the crystal structures of apo-CaM (1cfd) and CBP-bound-CaM (2bbm) show no changes in the surface exposure of Lys78 and Lys95. However, CBP binds to the region between Lys78 and Lys95 and likely sterically interferes with the ability of the Q2linker to form a crosslink between them (Fig. 2b). The results from the MBP and CaM qCLMS experiments show that the Qlinker approach can detect conformational changes in biosensors.

*Probing Rpb4/7-induced structural changes in pol II using Q2linkers*

We next evaluated the ability of Q2linkers to detect conformational changes in large protein complexes. Yeast pol II is a >0.5 MDa protein complex and a good model system to study structural changes in large complexes as the 12-subunit holo-pol II assumes a conformation that is distinct from that of the 10-subunit, core pol II complex, lacking the Rpb4/Rpb7 dimer[51-53]. In core pol II (1i3q), the "clamp" is in an "open" state, allowing formation of a straight channel for DNA template entry[51]. In the core pol II elongation complex (1nik) and the holo-pol II enzyme (5u5q), a massive movement of the "clamp", which rotates by about 30º with a maximum displacement >30 Å at external sites, results in the "closed" state[52, 54]. However, most of the clamp moves as a rigid body and the large structural movement is produced by conformational changes in five "switch" regions[54] (Fig. S4 magenta). The Rpb4/Rpb7 heterodimer binds to core pol II through interactions between Rpb6 (91-105) and Rpb1 (1440-1452) with Rpb7. *RPB4* is not required for viability while *RPB7* is essential for growth in yeast. We affinity purified holo-pol II from a yeast strain expressing FLAG-tagged Rpb3, and we purified pol II lacking Rpb4 from an *RPB4* deletion strain expressing FLAG-tagged Rpb2 (Fig. 3a). We crosslinked ΔRpb4 pol II and holo-pol II with C1q2 and C2q2, respectively in experiment I, and then performed a crosslinker swapping experiment in which we crosslinked ΔRpb4 pol II and holo-pol II with C2q2 and C1q2, respectively in experiment II (Fig. 3a). The samples were analyzed by mass spectrometry, and Comet database searching against the yeast proteome identified unmodified peptides and monolinks corresponding to 234 proteins (>99% probability). We used pLink2[55] and in-house designed Nexus[33] to identify crosslinks by searching a database composed of the sequences of the 12 pol II subunits, and then extracted the ion intensities for the 126 and 127 reporter ions from the identified spectra for their relative quantification. We then averaged the log2(126/127) ratios from each spectrum corresponding to a Qlinker-modified site or pair of crosslinked sites (Table S1), and plotted the ratios from experiment I on the x-axis and experiment II on the y-axis of the graphs shown in Figure 3b-d. Each green dot represents one unique site or pair of sites identified in both experiments. Each blue or yellow dot, represents a unique site or pair of sites identified only in experiment I or experiment II, respectively.

101 interlinks were identified in both experiment I and II, 47 interlinks were identified in experiment I only, and 116 interlinks were identified in experiment II only (Fig. 3b). However, among all of the interlinks that were only identified in one of the experiments, only 3 in experiment I and 4 in experiment II exhibit a greater than 2-fold intensity difference. This suggests that the inability to identity most of these crosslinked peptides in both experiments is mainly due to under

sampling during mass spectrometry analysis of the complex samples, rather than the absence of the crosslinked peptides in one of the experiments. This also highlights the importance of reliable methods for quantification in CLMS experiments involving complex samples, where under sampling of crosslinked peptides is exacerbated due to the increased complexity of the crosslinker-modified samples and the inefficiency of the crosslinking reaction. Considering all of the interlinks identified in both experiments, there is a rough anti-correlation between the ratios measured in experiment I and II ($R^2 = 0.5$) (Fig. 3b). We labelled all of the interlinks that exhibit more than a two-fold abundance difference in both experiments (Fig. 3b). The crosslinks between Rpb4:80-Rpb7:23, Rpb4:76-Rpb7:23 and Rpb1:2-Rpb7:29, labelled in green, were more abundant in the holo-pol II sample, as there is more Rpb4/7 present in this sample. Several interlinks and intralinks involving Rpb1:332, Rpb2:507, and Rpb1:1102, labelled in red, are also more abundant in holo-pol II (Figs. 3b and c). Rpb1:332 is located in the Switch 2 region (Rpb1:328-346), which is the main switch responsible for the conformational change from the "open" to "closed" state (Fig. 4a). The helix in the open state (cyan) flips out (gray) towards the cleft to contact DNA at the -2, -1 and +2 positions[54]. This conformational change is accompanied by stabilization of the Rpb2 forkloop 2 (Rpb2:503-508). The conformational change in Switch 2 brings Rpb1:332, Rpb1:1102 and Rpb2:507 closer to each other, allowing them to be crosslinked by Q2linkers (Fig. 4a).

The crosslinks between Rpb1:15-Rpb6:76 and Rpb1:15-Rpb6:72, labelled in blue, are influenced by a conformational change in Switch 5 (Rpb1:1431-1433), which undergoes a hinge-link bend[54] (Fig. 4b). In the core pol II structure, Rpb1:K15 forms a salt bond with Rpb1:D1442 (K15:NZ - D1442:OD2 distance of 3.8Å). In holo-pol II, Switch 5 bending pulls Rpb1:D1442 away from K15, breaking the salt bond (NZ-OD2 distance of 6.7Å). Even though the relative positions of Rpb1:K15, Rpb6:K76, and Rpb6:K72 do not change significantly in the holo- and core pol II structures (Fig. 4b), the increase in crosslink abundances involving these residues in the "open" state is likely attributed to the salt bond between K15 and D1442 in the "closed" state which blocks the NHS ester-based reaction between the epsilon amino group of K15 and the crosslinker. Not all of the crosslink abundance changes can be easily explained. Rpb1:1246 is located in a region that interacts with Rpb9 and Rpb2, and changes from a disordered region (1245-1254) in core pol II to a loop in holo-pol II[54]. The increased abundance of the Rpb1:1246-Rpb9:77 in holo-pol II might reflect this structural change. Rpb10:68 is very close to Rpb2:191, but there is no obvious difference in the location of the residues in both structures. Rpb1:129 is located at the N-terminus of Rpb1 (the major part of the clamp domain 1-346) that moves as a rigid body. A slight orientation difference of Rpb1:K129 may affect its ability to crosslink to Rpb5:K171. The crosslink between Rpb11:37-Rpb1:129 could be a false positive identification as these residues are located on opposite sides of pol II and the distance between their Cα atoms exceeds the theoretical crosslinking distance of Q2linkers.

287 intralinks were identified in both experiments, and 124 and 143 intralinks were identified only in experiment I and II, respectively (Fig 3c). The anti-correlation of log2(126/127) ratios for the intralinks identified in both experiments is not strong ($R^2 = 0.3$). All of the intralinks involving

Rpb4 are more abundant in holo-pol II as expected. In this study, the intralinks are less useful for probing conformational changes as the backbones of most subunits align pretty well in both complexes, and there are only small rearrangements for most residues. For example, the Rpb5:161-171 crosslink is enriched in the ΔRpb4-pol II sample, but there is little structural change for the region involved (Fig. 4c). Rpb5:K161 overlays perfectly in both structures and forms a salt bond with Rpb5:E172 with NZ-OE2 distances of 2.7Å and 2.8Å in the holo- and core pol II structures, respectively (Fig. 4c). Rpb5:K171 assumes slightly different orientations in both structures due to the formation of a helical structure between residues 171-175 in holo-pol II (Fig. 4c). Rpb5:K171 is slightly more stable in core pol II (cyan, b-factor 74) than in holo-pol II (gray, b-factor 216), which may account for its increased abundance in the ΔRpb4-pol II sample. The Rpb1:689-728 crosslink is enriched in the holo-pol II sample, while these two sites, located in the middle of two alpha-helices, are well-aligned (Fig. 4d). The distance between the two NZ atoms of the lysine residues is ~4 Å in core-pol II and ~6.6 Å in holo-pol II. The 4 Å distance in the ΔRpb4-pol II sample may limit the ability of the Q2linkers to react with both ε amines to form a crosslink between the two lysine side chains. 191 monolinked peptides were identified in both experiments and 10 and 11 monolinked peptides were only identified in experiment I or II, respectively. All of the Rpb4 and Rpb7 monolinks are enriched in the holo-pol II sample as expected. Some Rpb1 residues at the N-terminal clamp domain were also enriched in the holo-pol II sample (Fig. 3d). In general, and like intralinks, the monolinks are less informative in this study for probing conformational changes. Interestingly, some Rpb5 monolinks and intralinks are slightly enriched in the ΔRpb4-pol II sample, even though there is no large conformational or structural rearrangement associated with this region. Rpb5 interacts with the same domains of Rpb6 and Rpb1 that interact with Rpb4 and Rpb7. It is possible that Rpb4/7 stabilizes Rpb1, Rpb6, and Rpb5 and their interactions with one another so that the region is more rigid in holo-pol II and less reactive to Q2linkers.

## Discussion

In this paper, we describe a new approach for studying conformational and structural changes in proteins and protein complexes that is based on quantitative CLMS with a novel set of isobaric crosslinking reagents, called Q2linkers. Q2linkers are small and simple, amine-reactive molecules with an optimal extended distance of ~10 Å for CLMS. After crosslinking, the samples are combined for all subsequent steps in the analysis including enzymatic digestion, peptide fractionation/clean-up, and MS analysis, thus minimizing variations that may occur during these steps. The ability to avoid technical biases introduced during sample processing is especially important in qCLMS studies where conformational changes may be revealed by small but reproducible quantitative changes in crosslinking efficiency. The MS2-based reporter ion quantification is simple and compatible with most high-resolution mass spectrometers. The isobaric qCLMS technology can capture both conformational changes and structural re-arrangements in complex protein samples, and is well-suited to be adopted as a common strategy to study protein-protein interactions and conformational changes in large protein complexes under different conditions.

Isobaric crosslinkers[44, 45] are attractive for studying conformational and structural changes for a number of reasons. Unlike label-free approaches, they allow samples to be combined immediately after crosslinking and analyzed by MS together, thus avoiding potential quantification inaccuracies due to artefacts that can occur during sample processing such as different digestion efficiencies, differential sample losses during additional isotopic labeling and/or purification steps, or different extents of amino acid modifications (i.e., methionine oxidation, or N-terminal glutamate to pyroglutamate conversion). In addition, quantification of crosslinked peptides based on isotope labeling has been shown to be more accurate than label-free based methods[56] which likely is important for detecting subtle conformational changes. Finally, unlike approaches that employ isotopically heavy and light isotopes, isobaric crosslinkers do not increase the complexity of the MS1 spectra which may improve sensitivity and reproducibly. Our Qlinker design is one of the simplest for isobaric crosslinkers. While the basic structure could be expanded to iminodiacetic acid and iminodibutyric acid for specific applications, we think the iminopropionic-based Qlinker design may be the most useful for general practice.

One of the biggest challenges associated with CLMS technology is how to interpret and use the resulting distance restraints. The crosslinking results are often used in integrative modelling approaches to produce low to medium resolution structural models or for verification of cryo-EM structures[32, 34, 57]. It is even more challenging to use the crosslinking results to study conformational and structural changes. As we have shown in the ΔRpb4-pol II and holo-pol II experiments, many interlinks (32-53%) and intralinks (30-33%) were only identified in one of the experiments (Fig. 3). Most of the crosslinks identified in only one experiment showed no significant abundance change in the two samples, suggesting that their identification in only one experiment is likely due to under sampling during MS analysis of the complex samples rather than the presence of the crosslinked peptides in only one experiment. This, combined with potential technical issues associated with label free qCLMS analyses mentioned above, highlight some of the challenges associated with inferring conformational changes based on label free qCLMS data. Previously, we performed triplicate label-free qCLMS experiments to study structural changes in the histone octamer upon ISW2 interaction[58]. To alleviate issues due to under sampling, we only considered crosslinked peptides identified in at least two experiments for inferring conformational changes. Unfortunately, this strategy becomes less effective as sample complexity increases and under sampling issues are exacerbated. While computational approaches that align MS runs and match MS1 features across MS runs provide a way to alleviate under sampling issues[56, 59, 60], these approaches are not ideal. The Q2linker, isobaric crosslinker-based qCLMS strategy alleviates issues due to under sampling and sample handling that are often encountered with label-free approaches, and thus permits reliable identification and quantification of site-specific changes in crosslinker reactivity associated with structural differences in two samples.

In large protein complexes, such as the pol II complex, interlinks are more informative than intralinks and monolinks for inferring structural changes since they provide information about the proximities of the modified amino acids and their associated domains. MS-based technologies such as H/D exchange and active hydroxyl radical mapping can provide information about changes in surface exposure and residue accessibilities[30, 61-63]. However, in large protein complexes, changes in surface exposure and residue accessibility are often not associated with structural shifts. For example, compared to ΔRpb4-pol II, there is a large "clamp" domain movement in holo-pol II, but the clamp moves as a rigid body and this large structural rearrangement is controlled by small conformational changes of several "switches". The monolinks are not informative, as the accessibility of surface exposed lysine residues does not significantly change. In fact, even though there is a >30Å movement at the tip of the clamp, we did not detect significant ratio changes for the crosslinks that map to the clamp domain, because the conformation of the domain itself does not change significantly. Instead, we observed changes in the abundances of crosslinks involving residues associated with the switches that cause the large scale movement. Isobaric qCLMS technology provides a powerful way to characterize conformational changes in protein complexes that is difficult to achieve using currently available structural-MS- approaches. In addition, the Q2linker approach complements the information provided by high resolution structural approaches such as cryo-EM. Unlike cryo-EM or X-ray crystallography, qCLMS is performed in solution under near physiological conditions. Furthermore, CLMS can provide structural information about flexible or disordered regions which is often difficult to obtain by high resolution approaches.

Like most CLMS approaches, this isobaric qCLMS approach depends on the modification of specific chemical moieties and detection of the modified peptides. It may not capture conformational changes if there are no reactive groups near the region involved in the structural shift or the modified peptides are difficult to detect during MS analysis. We did not observe changes in crosslinker-modified peptide abundances derived from pol II in the presence and absence of α-amanitin. This implies that this strategy, like all MS-based strategies, can only be used for interpretation of positively identified crosslinks or monolinks. Under sampling is a common problem for MS analysis of complicated samples. Future development of Qlinker will involve the generation of affinity reagents that can enrich the crosslinker-modified peptides to improve sensitivity and quantification, and incorporation of multiple isobaric labels in the Qlinkers, like the TMT labels, so that multiple conditions and replicates can be analyzed simultaneously.

## Materials and Methods

### *Materials*

Di-tert-butyl 3,3'-Iminodipropionate was purchased from TCI America (Portland, OR). Bromoacetic acid 1-13C and 2-13C were purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA). The N,N'-Diisopropylcarbodiimide (DIC), N,N-Diisopropylethylamine (DIPEA), Dimethylformamide (DMF), Acetonitrile (ACN), Dichloromethane (DCM) 2,6-dimethylpiperidine, N,N,N',N'-Tetramethyl-O-(N-succinimidyl)uronium tetrafluoroborate

(TSTU), Human Brain Calmodulin, Bovine Serum Albumin (BSA) were purchased from Millipore Sigma (St. Louis, MO). Calmodulin binding peptide 1 (MLCK peptide) was purchased from Genscript Biotech Corp. (Piscataway, NJ). Bacterial maltose binding protein was purchased from Novus Biologicals (Littleton, CO). NGC chromatography system from Bio-Rad Laboratories (Hercules, CA) was used for reverse phase FPLC-C18 separation and SNAP ULTRA C18 flash cartridges (40g) were purchased from Biotage (Uppsala, Sweden).

*Design of isobaric Qlinkers*

We sought to develop a simple isobaric CLMS-based strategy, called Qlinker that permits probing conformational changes in proteins/protein complexes by reliably quantifying the relative abundances of crosslinker-modified peptides derived from samples of the protein/complex in different structural states. An important consideration when designing isobaric amine-reactive crosslinkers is that the m/z's of the reporter ions do not overlap the m/z's of other fragment ions which could skew quantification. Unlike iTRAQ or TMT modified peptides, which cannot generate unmodified $b1^+$ ions (because all peptide N-termini are labeled), the Qlinker-modified peptides, which are generated by enzymatic digestion after the crosslinking reaction is complete, have free N-terminal amines and will generate unmodified $b1^+$ ions. Since isoleucine and leucine residues generate $b1^+$ ions at 114.09 m/z and asparagine and aspartate residues generate $b1^+$ ions at 115.05 m/z and 116.03 m/z, respectively, we sought to incorporate a moiety into the Qlinker that generates reporter ions with m/z's that do not overlap with the m/z's of $b1^+$ and immonium ions. Thus, we decided to use 1-imino-2.6-dimethylpiperidin-1-ium as the reporter ion moiety which is the same as the moiety used in 126-134 TMT reagents (Fig. S1). We searched 83,242 mass spectra obtained from the analysis of a BS3 crosslinked sample of the 1 MDa yeast Mediator complex and only found 141 spectra (0.17%) containing 126.128 ions and 18 spectra (0.02%) containing 127.131 ions within a 40 ppm mass tolerance. These results indicated that the 126.128 and 127.131 reporter ions can be used for accurate quantification of Qlinker-modified peptides with little interference from fragment ions derived from unmodified or non-Qlinker-modified peptides.

Another important consideration in the design of the amine-reactive Qlinker is that it can react with ε-amines of lysine that are in close proximity. Iminodiacetic acid, iminodipropionic acid and iminodibutyric acid were considered as base structures. Due to the flexibility of lysine side chains, it was found that Cα-Cα distances of residues crosslinked with amine-reactive crosslinkers are more useful as distance restraints than the distances between crosslinked ε-amines. The BS3 crosslinker has a spacer arm of 11.4 Å when fully extended and can crosslink lysine residues whose Cα atoms are up to 30 Å apart[64]. The zero length crosslinker EDC/DMTMM can crosslink a lysine residue and an aspartate or a glutamate residue with Cα atoms up to 20 Å apart[65]. We reasoned that crosslinkers based on iminodiacetic acid or iminodipropionic acid with extended crosslinker spacer arms of 7.6 Å and 10.1 Å, respectively, would provide similar structural information, while the backbone of an iminodiacetic acid-based reagent would be more restricted than that of an iminopropionic acid-based reagent. At the same time, iminodibutyric acid quickly generates a dark

red color during activation. We decided to use iminodipropionic acid as the base structure for the new quantitative Qlinker (Fig. S1).

*Synthesis of bis(succinimidyl)-3,3'-{[(2,6-dimethylpiperidin-1-yl)acetyl]azanediyl}dipropanoic acid ("Q2linker")*

To synthesize the Qlinker, we first synthesized (2,6-dimethylpiperidin-1-yl) acetic acid and reacted it with di-tert-butyl-3,3'-iminodipropoinate ("**1**") but we obtained little desired product due to steric hindrance of dimethylpiperidine, resulting in difficulties activating the (2,6-dimethylpiperidin-1-yl) acetic acid. Then we used a peptoid synthesis strategy to have a one-pot synthesis of di-tert-butyl protected product "**2**" using the 1-$^{13}$C or 2-$^{13}$C bromoacetic acid (BAA) for the two isobaric crosslinkers (Fig. S1a). 0.5mL Di-tert-butyl 3,3'-Iminodipropionate ("**1**") (~ 3.2 M, 1 eq), 0.5mL DIC (~ 6.4 M, 2 eq) and 280 mg Bromoacetic acid 1-13C or 2-13C (1.25 eq) in 1 mL DMF were mixed for 2 h at RT. Then 10 mL DCM and 10 mL 0.1% trifluoroacetic acid (TFA) were used to extract product Di-tert-butyl-3,3'-[(bromoacetyl)azanediyl]dipropanoic acid in the DCM phase. 1.1 mL (~5 eq) dimethylpiperidine was added to the DCM extract and mixed at RT for 2 h and the solvent was then evaporated under vacuum. The product Di-tert-butyl-3,3'-{[(2,6-dimethylpiperidin-1-yl)acetyl]azanediyl}dipropanoic acid "**2**" was then dissolved in 50% ACN / 0.1% TFA and diluted to 20% ACN. The precipitates were removed by centrifugation and the supernatant was loaded onto a FPLC-RP-C18 cartridge at 5 mL/min and eluted with a 70 min gradient from 20% to 60% ACN. The product was monitored by 215 nm absorption and confirmed by mass spectrometry analysis using an LTQ (Thermo Scientific). The fractions containing the product "**2**" were then combined and evaporated by rotovap. 5 mL TFA was then added to the dried substance for 2 h and then evaporated by rotovap. The final product "**3**" of 3,3'-{[(2,6-dimethylpiperidin-1-yl)acetyl]azanediyl}dipropanoic acid was then dissolved in 0.1% TFA, loaded onto a FPLC-RP-C18 column, and eluted with a 50 min gradient from 0% to 20% ACN. Product "**3**" with MH+ of 316.34 was collected and was determined to be >95% pure. Product "3" was fractionated a second time using the same FPLC-RP-C18 conditions to yield ~50 mg of the final product at ~99% purity (~10% efficiency). In situ activation of Q2linker was achieved by mixing 0.25 M product "**3**" dissolved in dry DMF with an equal volume of 0.5 M TSTU in dry DMF and 1/10th volume of DIPEA at RT for one hour. The final stock concentration of Q2linkers was ~0.12 M and the crosslinkers can be stored at -20 ºC or -80 ºC for up to one month with little loss of activity. Hydrolysis of the activated crosslinker was observed upon longer storage, possibly due to trace-amounts of water in the DMF solution.

*General crosslinking protocol and workflow optimization*

To perform the crosslinking reaction, we typically used freshly-activated crosslinkers or brought the frozen crosslinkers to RT for 30 min before use. A typical crosslinking reaction is carried out in 50 mM HEPES buffer (pH 7.9) or 1X PBS (pH 7.5) with 20 to 100 μg total protein in a volume of 50 to 400 μL with ~2 mM (50X dilution from the stock) crosslinker. Nearly equal amounts of proteins were first crosslinked with one of the Q2linkers for one to two hours at RT and then 10 μL 1 M ammonium sulfate was added to quench each reaction for 10 min. The two crosslinked samples were then combined and vortexed to mix the sample. 10 μL 20 mg/mL SP3 magnetic beads was added to the sample and the sample was mixed well. ACN was then added to 70% and

the sample was incubated in a thermomixer at 60º C for 30 min with mixing. After collecting the beads on a magnetic stand, the beads were washed with 100% ACN and resuspended in 100 µL 8 M Urea, 50 mM choloroacetamide (CAA), 50 mM TCEP in 1 M ammonium bicarbonate buffer at 37 ºC for one hour. The beads were then diluted by adding 700 µL pure water and 1/10 (w/w) trypsin to digest the sample overnight at 37º C with rotation. The digested peptides were then purified using C18 micro-tips and dried for mass spectrometry analysis. Complex samples were fractionated by HPLC using in-house prepared microcapillary strong cation exchange columns (200 µm X 20 cm; SCX 3 µm, Sepax Technologies) at a flowrate of 2-3 µL/min. Peptides were eluted with 20 µl of Buffer A (10% ACN, 0.1% FA) containing 30%, 50%, 70%, and 100% Buffer B (800 mM ammonium formate, 20% ACN, pH 2.8), followed by 50 µl Buffer D (0.5 M ammonium acetate, 30% ACN). All fractions were dried in a Speed-vac, and resuspended in 0.1% TFA, 2% ACN.

We used the C1q2 and C2q2 reagents to crosslink BSA and analyzed the sample using a Thermo Oribtrap-Fusion mass spectrometer with an HCD collision energy of 28% or 30%. The monolinked peptides spectra were readily identified by Comet database searching using a differential modification of 297.1770 on lysine residues, but most spectra had very low or no reporter ion intensities. We then increased the HCD collision energy to 35%, 40% and 45% and observed higher reporter ion intensities, but collision energies above 35% resulted in poor fragmentation spectra and significantly reduced the numbers of identified peptides (data not shown). Synchronous precursor selection coupled with MS3 (SPS-MS3) technology can be used to quantify TMT labeled peptides because all of the b ions and lysine containing y ions are TMT labeled and can be selected for MS3 to generate reporter ions at high energy [66]. However, this technology cannot be used for Q2linker quantification because most fragment ions will not retain the reporter moiety and thus will not yield reporter ions during MS3. At the same time, the Q2linkers are only associated with modified lysine residues and may not produce observable daughter ions in the MS2 spectra that can be selected for MS3, especially for the crosslinked species. The Yates group reported that a stepped HCD strategy increased both the diversity of fragmentation ions and TMT reporter ion intensity[48]. Following their suggestion for TMT labeled peptides, we used stepped HCD settings of 24%, 30% and 36% and obtained MS2 spectra with much higher reporter ion intensities as well as better fragmentation across the peptide backbone. The average reporter ion intensity under this condition is about 60% ±25% of the highest peaks. Thus, this condition was used for all subsequent MS analysis of Q2linker-crosslinked samples.

*Purification of yeast pol I and pol II complexes*

We constructed yeast (*Saccharomyces cerevisiae)* strains carrying C-terminal His6-3XFLAG-His6-Ura3 (HFH) tags on *RPA2* and *RPB3* by swapping the tandem affinity purification (TAP) tag for the HFH tag in strains carrying C-terminal TAP tags on *RPA2* and *RPB3*. The *RPB2-HFH* strain was generated by transforming a Δ*rpb4* strain with a PCR product containing the HFH tag with 40 bps of sequence flanking the stop codon of *RPB2* gene. For each complex purification, we normally grew 6 L of the appropriate yeast strains in YPD media overnight to OD600 11-13. The cells were harvested by centrifugation and frozen in liquid $N_2$. After evaporation of the liquid $N_2$, the cell pellets were ground to a fine powder in a coffee grinder. 40 mL lysis buffer (50 mM HEPES, pH 7.9, 400 mM ammonium sulfate, 10 mM MgSO4, 1 mM EDTA, 20% glycerol) with

protease inhibitors was then added to the fine powder and the mixture was stirred at 4º C for one hour. The cell lysate was then sonicated using a Superhorn sonicator with output 9 in an ice-water bath using 1 minute cycles of 30 seconds on, 30 seconds off, over 10 minutes. The lysate was centrifuged at 20,000 × g for 1 hour and the supernatant was then mixed with 2 ml anti-FLAG M2 affinity agarose (Sigma-Aldrich) overnight at 4 ºC with rotation. The beads were then collected in a column and washed with 40 mL lysis buffer twice, 40 mL 2X PBS buffer, and 40 mL 1X PBS with 0.1% NP-40. The complexes were eluted with 3X FLAG peptide (Sigma-Aldrich) at 0.4 mg/ml in 1X PBS. The complexes were then concentrated by repeated centrifugation and dilution with 1X PBS (normally three times) in Amicon Ultra-4 devices (100K cutoff, Millipore) to reduce the concentration of the 3X FLAG peptide. After a final centrifugation step, protein concentration was determined by Qubit protein assay (Thermo Fisher Scientific). We usually isolated ~200 to 300 µg protein from a 6 L culture. Higher yields of pol II were obtained from the *ΔRPB4* strain. We checked the purity of the sample and subunit composition by Coomassie stained SDS-PAGE.

*Mass spectrometry analysis and data processing*

Peptides were analyzed by electrospray ionization microcapillary reverse phase HPLC with a column (75 µm × 270 mm) packed with ReproSil-Pur C18AQ (3-µm 120Å resin; Dr. Maisch, Baden-Würtemburg, Germany) on a Thermo Scientific Fusion with HCD fragmentation and serial MS events that included one FTMS1 event at 30,000 resolution followed by FTMS2 events at 15,000 resolution. Other instrument settings included: MS1 scan range (m/z): 400-1500; cycle time 3 sec; Charge states 3-8; Filters MIPS on, relax restriction = true; Dynamic exclusion enabled: repeat count 1, exclusion duration 30s; Filter IntensityThreshold, signal intensity 50000; Isolation mode, quadrupole; Isolation window 3Da; Activation Type: HCD; Collision Energy Mode: Stepped; HCD Collision Energy (%): 24, 30, 36; isolation width: 3Da; AGC target 500,000, Max injection time 200ms. HPLC uses an 80 min gradient from 10% ACN to 40% ACN.

The RAW files were converted to mzXML files by Rawconverter[67] . For normal peptide and monolinked peptide searches, we used the Trans-Proteomics Pipeline (TPP)/Comet searches (http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP) with static modification on cysteines (+57. 0215 Da) and differential modifications on methionines (+15.9949 Da) and lysines (+297.166962 Da and +279.166397 Da). Spectra identified as peptides with lysine modification(s) with PeptideProphet probability >95% were output as spectra for monolinked peptides. For crosslinked peptide searches, we used two crosslink database searching algorithms: pLink2[55] and an in-house designed Nexus[33] with Q2linker mass of 279.1664 Da against a database containing only yeast pol I or pol II protein sequences. Other searching parameters include: precursor monoisotopic mass tolerance: ±20 ppm; fragment mass tolerance: ±20 ppm; up to three miscleavages; static modification on cysteines (+57. 0215 Da); differential oxidation modification on methionines (+15. 9949 Da), peptide N-terminal glutamic acid (-18.0106 Da) or N-terminal glutamine (-17.0265 Da); Q2linker modification on lysines (+297.166962 Da). After performing the pLink and the Nexus analyses, the search results were combined and each spectrum was manually evaluated for the quality of the match to each peptide using the COMET/Lorikeet Spectrum Viewer (TPP). The crosslinked peptides are considered confidently identified if at least

4 consecutive b or y ions for each peptide are observed and the majority of the observed ions are accounted for. Search results that did not meet these criteria were removed. The spectra that passed our evaluation are summarized in Table S1 and are uploaded into ProXL[68] for viewing and data analysis. All of the data including the spectra, linkages and structural analyses can be visualized at https://www.yeastrc.org/proxl_public/viewProject.do?project_id=634. The raw files are deposited at protomeXchange: PXD035939 ( **Username:** reviewer_pxd035939@ebi.ac.uk **Password:** XooGYvtL).

*Q2linker quantification*

A perl script is used to extract the 126 and 127 reporter ion intensities for the identified monolinked- or crosslinked-spectra from mzXML files within a mass tolerance of 0.005 Da (~ 40 ppm) of the theoretical 1+ mass of 126.127726 and 127.131081 for the 126 and 127 reporter ions, respectively. These observed intensities are designated **I**126 and **I**127, respectively. Like TMT quantification, these observed intensities need to be adjusted based on the natural distribution of monoisotopic elements and the 99% purity of the heavy 13C element. It is relatively simple for the Q2linkers since only two channels need to be considered. We corrected the reporter ion intensities (**A**126 and **A**127) by solving the equations: **A**126 * 0.90754 + **A**127 * 0.0 1 = **I**126 and **A**126 * 0.09246 + **A**127 * 0.90724 = **I**127. The final intensities **A**126 = (**I**126 -**I**127*0.01102) / 0.906521 and **A**127 = (**I**127 - **I**126*0.10188) / 0.906221 were used to calculate the log2 ratios of the126 and 127 reporter ions. If multiple spectra were identified for a crosslinker-modified site, we used the average log2 ratio of all the identified spectra corresponding to that site.

## Statistics and Reproducibility

Average and boxplot in R were used to generate Figure 1d and e. Crosslinker swapping experiments were performed for the experiments presented in Figures 2a and 3.

## Data availability statement

The raw MS files for the data presented in figures 3 and 4 are deposited at proteomeXchange: PXD035939 ( **Username:** reviewer_pxd035939@ebi.ac.uk **Password:** XooGYvtL). The datasets generated during and/or analysed during the current study are available from the corresponding author on request.

## Conflict of interest statement

The authors declare no competing interests.

## Acknowledgements

Funding support was provided by the NIH (R01GM136974) to J.R.

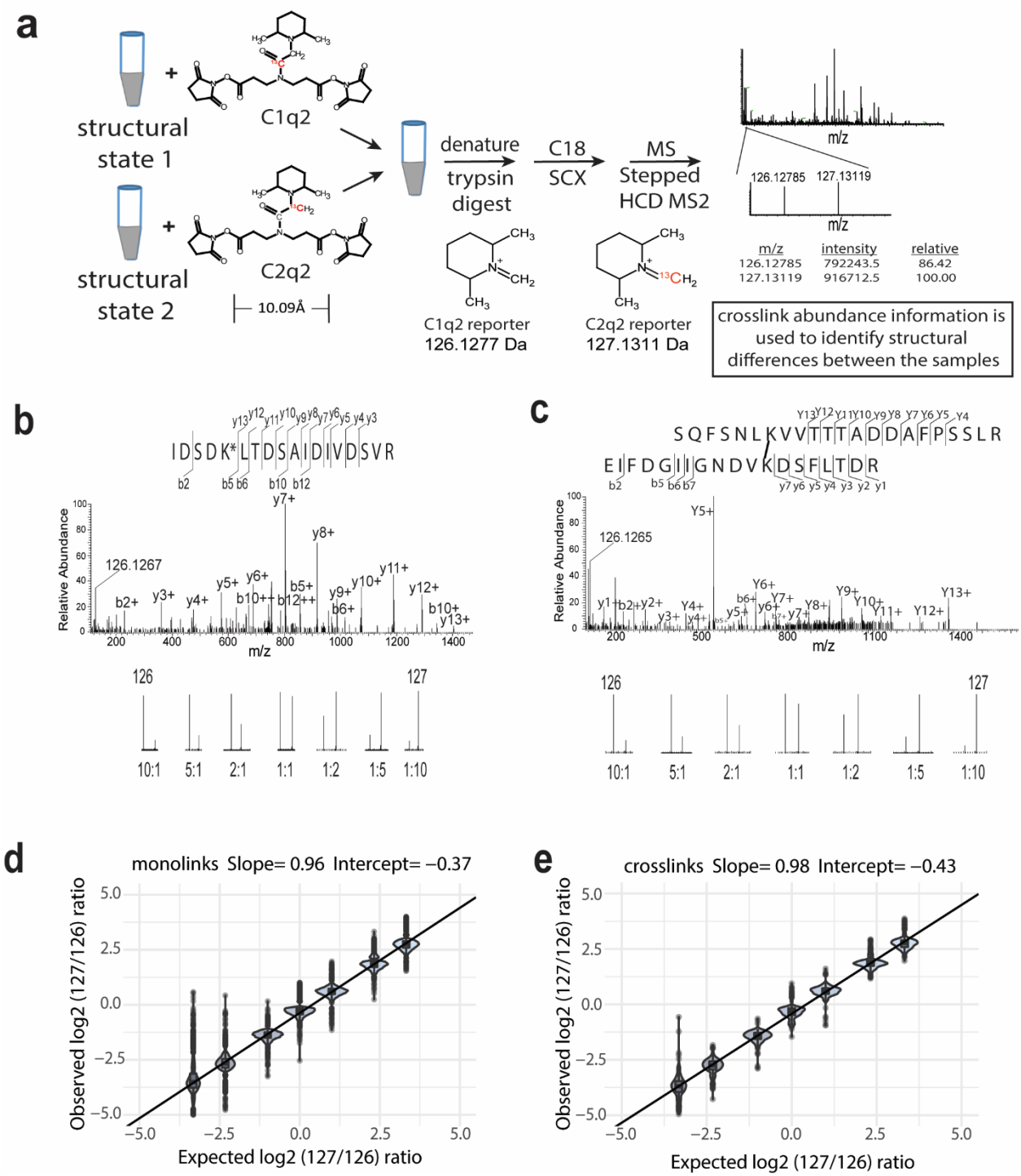


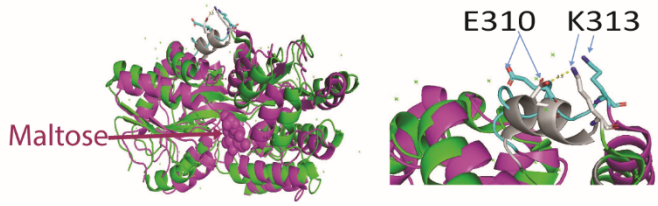

**Figure. 1. The isobaric Qlinker quantitative CLMS approach for studying conformational and structural changes in proteins and protein complexes. a)** The structure of Q2linkers and the general scheme for qCLMS using Q2linkers. The $^{13}C$ atom in C1q2 or C2q2 is indicated by red font. **b-e**: Experiment to evaluate the ability of Q2linkers to quantify the relative abundances of crosslinks and monolinks derived from Q2linker modification of affinity purified pol I at designated ratios. Example spectra of a monolinked peptide **(b)** and a crosslinked peptide **(c)** with

the observed reporter ion intensities at each mixing ratio. **d)** Observed vs. expected log2 (127/126) reporter ion ratios for monolinks. **e)** Observed vs. expected log2 (127/126) reporter ion ratios for crosslinks. Boxplot in R was used to create the graphs in d) and e). Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

**a**

K.SYEEELAK*DPR.L

| | label 126 | label 127 | log2(127/126) | abundance ratio with/without maltose |
|---|---|---|---|---|
| maltose | - | + | -2.34 | 5.06 |
| maltose | + | - | 2.40 | 5.35 |




**b**

| peptide | type | position | Log2(127/126) | change(fold) |
|---|---|---|---|---|
| R.VFDK*DGNGYISAAELR.H | monolink | K95 | 0.91 | 1.88 |
| R.KMK*DTDSEEEIREAF.R | monolink | K78 | 0.75 | 1.68 |
| R.VFDK*DGNGYISAAELR.H --- R.KMK*DTDSEEEIREAF.R | crosslink | K78---K95 | -2.52 | -5.73 |

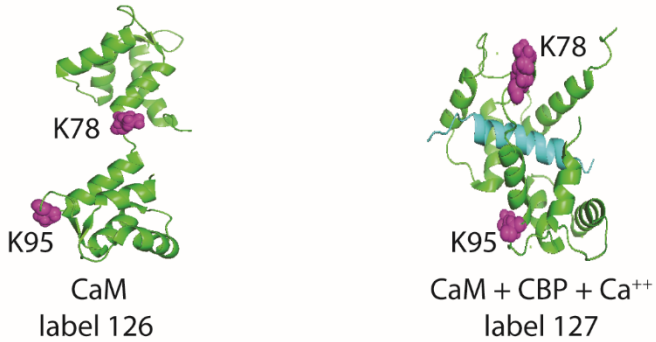



**Figure. 2. Q2linkers detect conformational changes in protein biosensors. a**) Quantification of a monolinked peptide from maltose binding protein (MBP) with or without maltose. Structures of MBP in the open, ligand-free conformation (green, 1mpb.pdb) and the closed, maltose-bound conformation (brown, 1n3w.pdb) are shown on the left. A close-up view of the "balancing interface" is shown on the right. In the closed conformation, the sequence between amino acids 301 to 312 forms an alpha helix (gray) and K313 forms a salt bond with E310. The helix and the salt bond are disrupted in the open conformation (cyan). A crosslinker swapping experiment was performed on the same preparation of MBP. **b)** Quantification of monolinked and crosslinked peptides involving K78 and K95 from apo-CaM and CaM with $Ca^{2+}$ and CBP. Structures of apo-CaM (left) and CaM + CBP (blue) + $Ca^{2+}$ (right) are shown with key lysine residues space filled and magenta.

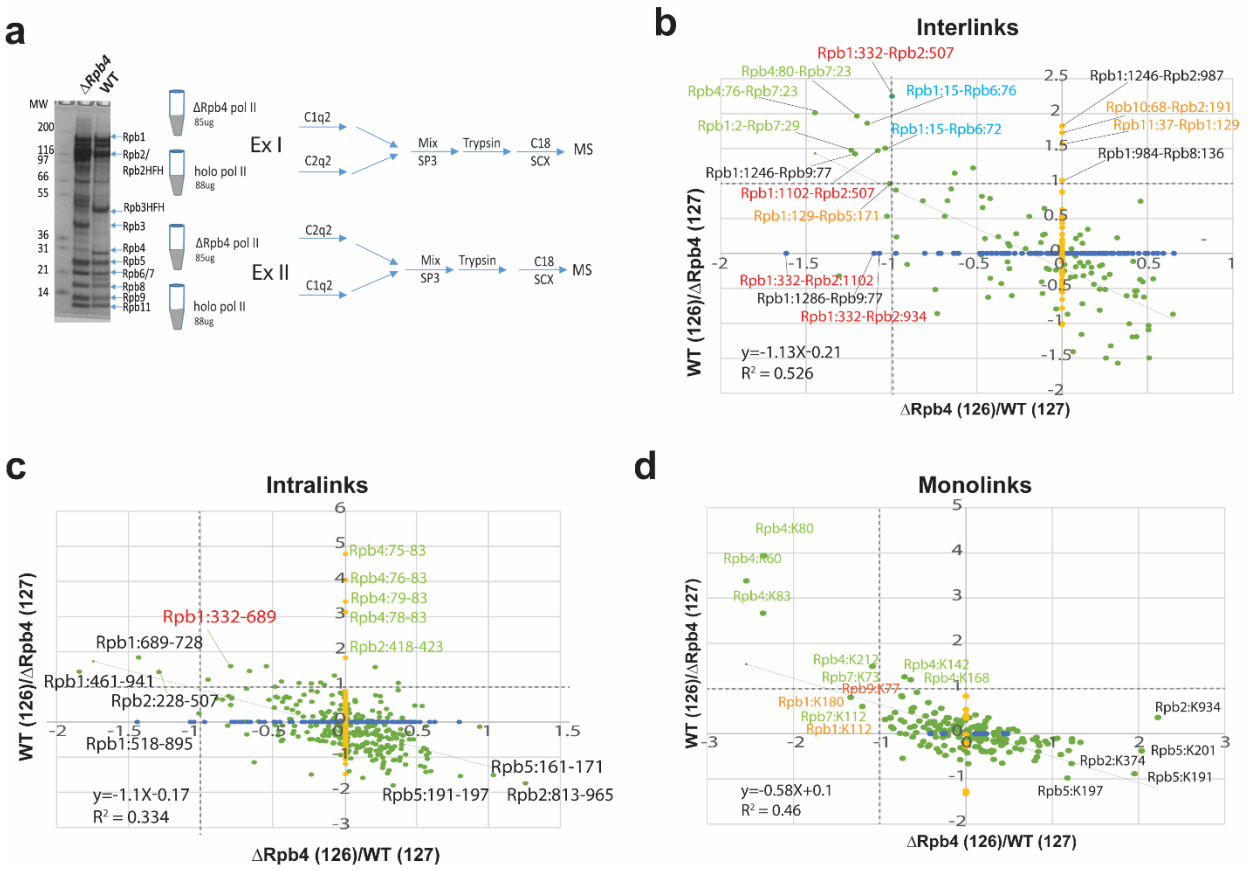



**Figure. 3. Q2linkers detect conformational changes in large protein complexes a)** The SDS-PAGE gel of affinity purified pol II from WT and ΔRpb4 strains and experimental design for studying conformational changes in RNA polymerase II due to deletion of Rpb4. **b-d)** The log2(126/127) ratio comparisons in experiment I and II for holo-Pol II and ΔRpb4-pol II. x-axis is ΔRpb4-pol II (126)/holo-pol II (127). y-axis is holo-pol II (126)/ΔRpb4-pol II (127). Enrichment in the holo-pol II sample (labelled with C2q2 in experiment I) will have a smaller log2(126/127) ratio on the x-axis; enrichment in the holo-pol II (labelled with C1q2 in experiment II) will have a higher log2(126/127) ratio on the y-axis. Each green dot corresponds to one unique pair of crosslinking sites identified in both experiments. Each blue dot on the x-axis corresponds to one unique pair of crosslinking sites identified only in experiment I and each yellow dot on the y-axis corresponds to one unique pair of crosslinking sites identified only in experiment II. Only the green dots are used for the linear regression analysis. **b)** interlinks, **c)** intralinks and **d)** monolinks. Dashed lines indicate log2 ratio = 1. The same preparation of each protein complex was used in a crosslinker swapping experiment.

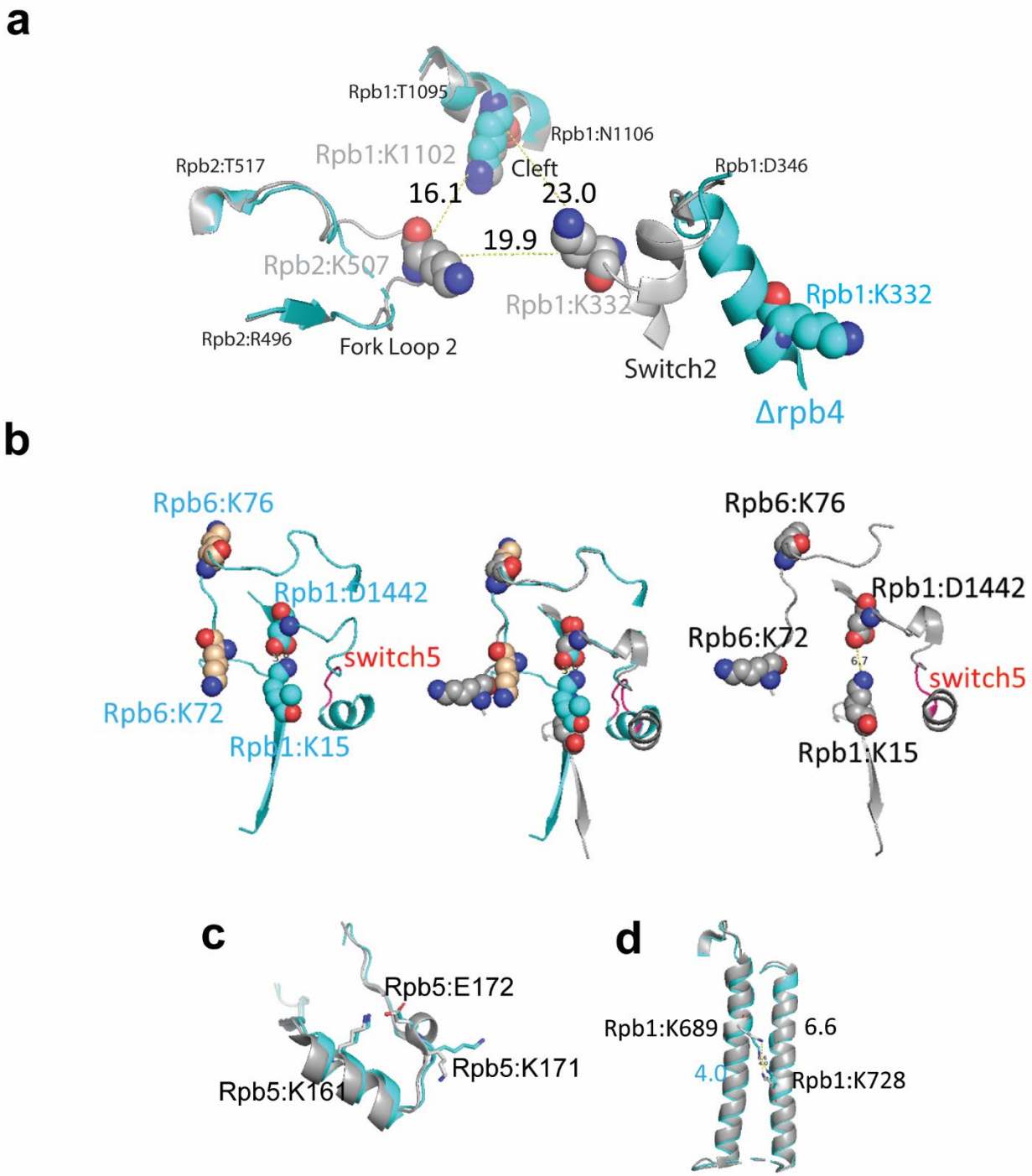



**Figure 4. Q2linkers detect conformational changes in pol II complexes. a)** Switch 2 (Rpb1:328-346) conformational changes in ΔRpb4-pol II (cyan) and holo-pol II (gray). The crosslinked lysine residues are shown as spheres. Cα– Cα distances between crosslinked lysines are indicated **b)** Structural comparison of crosslinks involving switch 5 (red, Rpb1:1431-1433) for ΔRpb4-pol II (cyan, left), holo-pol II (gray, right) and the merged structures (middle). **c)** Structural

comparison of the region of Rpb5 involving the crosslink between K161 and K171 in ΔRpb4-pol II and holo-pol II. A salt bond (not shown) is formed between K161 and E172 in both structures. **d)** Structure of Rpb1 near the crosslink between Rpb1:K689 and Rpb1:K728 in ΔRpb4-pol II and holo-pol II. In all structures, ΔRpb4-pol II is cyan and holo-pol II is gray. NZ–NZ distances between crosslinked lysines are indicated.

## References


1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
2. Sannigrahi, A., De, N. & Chattopadhyay, K. The bright and dark sides of protein conformational switches and the unifying forces of infections. *Commun Biol* **3**, 382 (2020).
3. Sicoli, G., Vezin, H., Ledolter, K., Kress, T. & Kurzbach, D. Conformational tuning of a DNA-bound transcription factor. *Nucleic Acids Res* **47**, 5429-5435 (2019).
4. Khan, S.H. & Kumar, R. An overview of the importance of conformational flexibility in gene regulation by the transcription factors. *J Biophys* **2009**, 210485 (2009).
5. Darling, A.L. & Uversky, V.N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* **9**, 158 (2018).
6. Mashtalir, N. *et al.* A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell* **183**, 802-817 e824 (2020).
7. Prabantu, V.M., Naveenkumar, N. & Srinivasan, N. Influence of Disease-Causing Mutations on Protein Structural Networks. *Front Mol Biosci* **7**, 620554 (2020).
8. Carrell, R.W. & Gooptu, B. Conformational changes and disease--serpins, prions and Alzheimer's. *Curr Opin Struct Biol* **8**, 799-809 (1998).
9. Vos, S.M. *et al.* Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612 (2018).
10. Catterall, W.A., Wisedchaisri, G. & Zheng, N. The conformational cycle of a prototypical voltage-gated sodium channel. *Nat Chem Biol* **16**, 1314-1320 (2020).
11. Pilla, K.B., Leman, J.K., Otting, G. & Huber, T. Capturing conformational States in proteins using sparse paramagnetic NMR data. *PLoS One* **10**, e0127053 (2015).
12. Chen, J. *et al.* Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication-transcription complex. *Nat Struct Mol Biol* **29**, 250-260 (2022).
13. Burger, V.M., Arenas, D.J. & Stultz, C.M. A Structure-free Method for Quantifying Conformational Flexibility in proteins. *Sci Rep* **6**, 29040 (2016).
14. Nadeau, O.W. & Carlson, G.M. A review of methods used for identifying structural changes in a large protein complex. *Methods Mol Biol* **796**, 117-132 (2012).
15. Sengupta, I. & Udgaonkar, J. Monitoring site-specific conformational changes in real-time reveals a misfolding mechanism of the prion protein. *Elife* **8** (2019).
16. Rinas, A., Mali, V.S., Espino, J.A. & Jones, L.M. Development of a Microflow System for In-Cell Footprinting Coupled with Mass Spectrometry. *Anal Chem* **88**, 10052-10058 (2016).
17. Wang, L. & Chance, M.R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol Cell Proteomics* **16**, 706-716 (2017).
18. Sala, D., Del Alamo, D., McHaourab, H.S. & Meiler, J. Modeling of protein conformational changes with Rosetta guided by limited experimental data. *Structure* (2022).
19. Patel, A.B. *et al.* Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362** (2018).
20. Akey, C.W. *et al.* Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* **185**, 361-378 e325 (2022).
21. Liu, X.R., Zhang, M.M. & Gross, M.L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem Rev* **120**, 4355-4454 (2020).

22. Haymond, A. *et al.* Protein painting, an optimized MS-based technique, reveals functionally relevant interfaces of the PD-1/PD-L1 complex and the YAP2/ZO-1 complex. *J Biol Chem* **294**, 11180-11198 (2019).
23. Luchini, A., Espina, V. & Liotta, L.A. Protein painting reveals solvent-excluded drug targets hidden within native protein-protein interfaces. *Nat Commun* **5**, 4413 (2014).
24. Bamberger, C. *et al.* Protein Footprinting via Covalent Protein Painting Reveals Structural Changes of the Proteome in Alzheimer's Disease. *J Proteome Res* **20**, 2762-2771 (2021).
25. Liu, H. *et al.* Mass spectrometry footprinting reveals the structural rearrangements of cyanobacterial orange carotenoid protein upon light activation. *Biochim Biophys Acta* **1837**, 1955-1963 (2014).
26. Zhou, Y. & Vachet, R.W. Covalent labeling with isotopically encoded reagents for faster structural analysis of proteins by mass spectrometry. *Anal Chem* **85**, 9664-9670 (2013).
27. Kahsai, A.W. *et al.* Multiple ligand-specific conformations of the beta2-adrenergic receptor. *Nat Chem Biol* **7**, 692-700 (2011).
28. Hambly, D.M. & Gross, M.L. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J Am Soc Mass Spectrom* **16**, 2057-2063 (2005).
29. Espino, J.A., Mali, V.S. & Jones, L.M. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Anal Chem* **87**, 7971-7978 (2015).
30. McKenzie-Coe, A., Montes, N.S. & Jones, L.M. Hydroxyl Radical Protein Footprinting: A Mass Spectrometry-Based Structural Method for Studying the Higher Order Structure of Proteins. *Chem Rev* **122**, 7532-7561 (2022).
31. Chen, Z.A. *et al.* Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *Embo J* **29**, 717-726 (2010).
32. Luo, J. *et al.* Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. *Mol Cell* **59**, 794-806 (2015).
33. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288 e1220 (2018).
34. Abdella, R. *et al.* Structure of the human Mediator-bound transcription preinitiation complex. *Science* **372**, 52-56 (2021).
35. Yu, C. *et al.* Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes. *Anal Chem* **88**, 10301-10308 (2016).
36. Ruwolt, M. *et al.* Optimized TMT-Based Quantitative Cross-Linking Mass Spectrometry Strategy for Large-Scale Interactomic Studies. *Anal Chem* **94**, 5265-5272 (2022).
37. Chen, Z.A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat Protoc* **14**, 171-201 (2019).
38. Mendes, M.L. *et al.* An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* **15**, e8994 (2019).
39. Muller, D.R. *et al.* Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal Chem* **73**, 1927-1934 (2001).
40. Petrotchenko, E.V., Olkhovik, V.K. & Borchers, C.H. Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol Cell Proteomics* **4**, 1167-1179 (2005).
41. Ihling, C.H. *et al.* The Isotope-Labeled, MS-Cleavable Cross-Linker Disuccinimidyl Dibutyric Urea for Improved Cross-Linking/Mass Spectrometry Studies. *J Am Soc Mass Spectrom* **31**, 183-189 (2020).
42. Makepeace, K.A.T. *et al.* Improving Identification of In-organello Protein-Protein Interactions Using an Affinity-enrichable, Isotopically Coded, and Mass Spectrometry-cleavable Chemical Crosslinker. *Mol Cell Proteomics* **19**, 624-639 (2020).

43. Chavez, J.D. *et al.* Quantitative interactome analysis reveals a chemoresistant edgotype. *Nat Commun* **6**, 7928 (2015).
44. Chavez, J.D., Keller, A., Mohr, J.P. & Bruce, J.E. Isobaric Quantitative Protein Interaction Reporter Technology for Comparative Interactome Studies. *Anal Chem* **92**, 14094-14102 (2020).
45. Chavez, J.D., Keller, A., Wippel, H.H., Mohr, J.P. & Bruce, J.E. Multiplexed Cross-Linking with Isobaric Quantitative Protein Interaction Reporter Technology. *Anal Chem* **93**, 16759-16768 (2021).
46. Wippel, H.H., Chavez, J.D., Tang, X. & Bruce, J.E. Quantitative interactome analysis with chemical cross-linking and mass spectrometry. *Curr Opin Chem Biol* **66**, 102076 (2022).
47. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
48. Diedrich, J.K., Pinto, A.F. & Yates, J.R., 3rd Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. *J Am Soc Mass Spectrom* **24**, 1690-1699 (2013).
49. Telmer, P.G. & Shilton, B.H. Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *J Biol Chem* **278**, 34555-34567 (2003).
50. Meister, G.E. & Joshi, N.S. An engineered calmodulin-based allosteric switch for Peptide biosensing. *Chembiochem* **14**, 1460-1467 (2013).
51. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-1876 (2001).
52. Bushnell, D.A. & Kornberg, R.D. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**, 6969-6973 (2003).
53. Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci Rep* **7**, 44628 (2017).
54. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**, 1876-1882 (2001).
55. Chen, Z.L. *et al.* A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* **10**, 3404 (2019).
56. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Methods* **12**, 1185-1190 (2015).
57. Patel, A.B. *et al.* Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *Elife* **8** (2019).
58. Hada, A. *et al.* Histone Octamer Structure Is Altered Early in ISW2 ATP-Dependent Nucleosome Remodeling. *Cell Rep* **28**, 282-294 e286 (2019).
59. Valot, B., Langella, O., Nano, E. & Zivy, M. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572-3577 (2011).
60. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
61. Zheng, J., Strutzenberg, T., Pascal, B.D. & Griffin, P.R. Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr Opin Struct Biol* **58**, 305-313 (2019).
62. Tsirigotaki, A., Elzen, R.V., Veken, P.V., Lambeir, A.M. & Economou, A. Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry. *Sci Rep* **7**, 2456 (2017).
63. Huang, W., Ravikumar, K.M., Chance, M.R. & Yang, S. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. *Biophys J* **108**, 107-115 (2015).
64. Merkley, E.D. *et al.* Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci* **23**, 747-759 (2014).

65. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci U S A* **111**, 9455-9460 (2014).
66. McAlister, G.C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **86**, 7150-7158 (2014).
67. He, L., Diedrich, J., Chu, Y.Y. & Yates, J.R., 3rd Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Anal Chem* **87**, 11361-11367 (2015).
68. Riffle, M., Jaschob, D., Zelter, A. & Davis, T.N. ProXL (Protein Cross-Linking Database): A Platform for Analysis, Visualization, and Sharing of Protein Cross-Linking Mass Spectrometry Data. *J Proteome Res* **15**, 2863-2870 (2016).